Mapping Quantitative Trait Loci and Assessing the Prospects of Genomic Prediction for Resistance to Goss's Wilt of Maize

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Amritpal Singh

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Aaron J. Lorenz

July 2017

Acknowledgements

I would like to thank my advisor Dr. Aaron J. Lorenz for giving me the opportunity to conduct this research. His contributions to the design and implementation of this research, guidance, support and encouragement throughout my time both at the University of Nebraska and the University of Minnesota helped me finish this dissertation. I appreciate him for always bringing up new ideas which were helpful in my personal and professional development. I am grateful to my dissertation committee chair Dr. Candice Hirsch for her valuable comments, suggestions and improvement of this dissertation specifically RNA sequencing experiment design. I also thank my committee members Dr. Rex Bernardo and Dr. Dean Malvick for providing expert guidance throughout the course of this research.

I would like to thank my previous committee members at the University of Nebraska including Dr. P. Stephen Baenziger, Dr. Jennifer Clarke, Dr. Tamra Jackson-Ziems, and Dr. James Alfano for their guidance during my time in Lincoln.

Aaron Andersen, previous student in Lorenz lab who worked on Goss's wilt before me helped in smooth transitioning and setting up initial experiments. I appreciate him for his help. Thanks are due to all other members of my lab at UNL including Diego Jarquin, Jon Luetchens, Nonoy Bandillo, Liakat Ali, and Collin Lamkey for their help and companionship. I also appreciate the help received from undergraduate students.

I also thank my fellow lab mates at U of MN including D.C. Kadam, Erin Gilbert, Siddhi Bhusal, Benjamin Campbell, Austin Dobbels, Becky Zhong, and Ryan Merry with whom I have had the pleasure to work during my time at the U of MN.

This project involved collaborations with other labs whose expertise helped in accomplishing specific components of the research. I want to thank Jackson lab members: Brad Tarnish, Jae Brungardt, Cody Kaarstad for their help in field inoculum preparation; Alfano lab members: Fan Yang, and Guangyong Li for molecular plant pathology related support specifically lab inoculations and preparation of RNA for sequencing; Alex Brohammer from Hirsch lab for help in RNA sequencing data analysis.

My studies were made possible by the funding support from Dow AgroSciences. I appreciate valuable feedback from Dow AgroSciences employees especially Dr. John Zheng, and Dr. Andrew Hopkins.

I would also like to thank the faculty, staff and my fellow graduate students in the Departments of Agronomy and Horticulture at UNL, and Agronomy and Plant Genetics at U of MN for making my experience rewarding and memorable. Finally, I thank my family and friends for their encouragement and unwavering support during my time in the graduate school. Special thanks to Rajesh, Manmeet, Kamaldeep, and Jatinder for their advice and support throughout the process.

This work was completed utilizing the Holland Computing Center of the UNL and Minnesota Supercomputing Institute of the U of MN.

Dedication

Dedicated to my mother

Abstract

Goss's wilt is a bacterial disease of maize caused by the Gram-positive bacterium *Clavibacter michiganensis* subsp. *nebraskensis* (*Cmn*). Goss's wilt was discovered for the first time in South Central Nebraska in 1969. Following its discovery, the disease spread to the neighboring states over the next decade. Maize germplasm was screened for resistance to Goss's wilt, and possibly due to the deployment of partially resistant hybrids, Goss's wilt did not cause any significant damage during the 1980s and 1990s. However, Goss's wilt re-emerged around 2006 and has been spreading to major maize growing areas in the United States and Canada. It is important to understand the genetic basis of resistance to Goss's wilt to devise strategies for breeding resistance into maize hybrids. The main objectives of this dissertation were to (i) map quantitative trait loci (QTL) for resistance to Goss's wilt using linkage mapping, joint linkage mapping, and genome-wide association mapping; (ii) identify differentially expressed genes in resistant and susceptible inbred lines in response to *Cmn* using RNA-seq; and (iii) to explore the prospects of genomic prediction of resistance to Goss's wilt. Three bi-parental linkage mapping families including B73 x Oh43, B73 x HP301, and B73 x P39 that were evaluated for Goss's wilt were used for joint linkage and linkage mapping. Eleven QTL were detected for resistance to Goss's wilt on chromosomes 1, 2, 3, 4, 5, and 10 through joint linkage mapping. Linkage mapping in each of the three families identified nine, six, and four QTL in the families B73 × Oh43, B73 × HP301, and B73 × P39, respectively. Genome-wide association analysis conducted using a diversity panel of 555 maize inbred lines and 450 recombinant inbred lines (RILs) from three bi-parental mapping

populations found three SNPs in the diversity panel and 10 SNPs in the combined dataset of diversity panel and RILs that were associated with Goss's wilt resistance. Two modules of correlated genes were discovered that showed differential regulation in response to *Cmn* between resistant (N551) and susceptible (B14A) inbred lines using a weighted gene co-expression network analysis. Gene ontology analysis revealed that the genes inside one of the modules were enriched in defense related functions.

Genomic prediction of Goss's wilt resistance was conducted on the data obtained from bi-parental families and the diversity panel. Highest predictive ability of 0.56 and 0.64 was achieved in the diversity panel and B73 x Oh43 population respectively. Effect of training population size, composition, and adding diverse lines to training population on predictive ability was also assessed. Results indicated that predictive ability is not highly benefited when training population is designed by adding equal number of lines from each of the three families. Adding diverse lines to the training population lead to minor changes in predictive ability.

Overall, the results improved our understanding of the genetic architecture of Goss's wilt resistance and showed that the resistance to Goss's wilt is a complex trait, controlled by small effect QTL.

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction and Review of Literature

1.1 Introduction

Maize is one of the most important cereal crops, as it ranks first among the cereals in terms of production and third in value after wheat and rice in the world. In 2014, maize was grown on 182 million hectares leading to 1.04 billion tonnes of maize production worldwide (FAOSTAT, assessed Nov. 29, 2016). In the United States, total maize grain production is estimated to be 361 million tonnes, making it the number one crop in terms of production (USDA-NASS). In terms of dollar value, maize grain contributed 49 billion US dollars to the United States economy in 2015 (USDA-NASS).

There has been a tremendous increase in maize yields in the United States over the last 70 years after the introduction of hybrid maize breeding in the form of double cross hybrids (1940 - 1960) and then single crosses (1960 - 2016). Although maize production has increased tremendously in the United States due to continuous efforts of plant breeders and agronomists, significant challenges still remain in order to drive maize production forward to feed an ever growing world population. Emerging insects, pests, and diseases continuously limit maize production and need to be addressed in order to sustain maize yield. This chapter is primarily focused on maize diseases, particularly Goss's wilt which is an emerging problematic disease in the Midwestern Corn Belt.

Despite efforts to protection against crop pests, losses in maize yield due to pests are substantial. From 2001-03, the loss in maize yield due to pathogens was estimated to be 8.5% (Oerke, 2006). Specific plant disease epidemics can cause significant economic damage. For example, the southern corn leaf blight (SLB) epidemic of 1970-71 is

important in the history of maize. In the Corn Belt states including Iowa and Illinois, 50-100% yield loss occurred in some fields and the average loss was estimated to be 20-30% due to SLB (Ullstrup, 1972), totaling to nearly one billion dollars lost (Ullstrup, 1972). The impact of the SLB epidemic on the United Sates agriculture was so severe that a national conference on leaf blight information in 1970 was attended by the United States President Richard Nixon. Another recent maize disease epidemic occurred in Sub-Saharan Africa. A disease of maize known as Maize Lethal Necrosis, which emerged in 2011 in Kenya, has now spread to seven Sub-Saharan African countries (Mahuku et al. 2015). During 2012, in Kenya alone maize lethal necrosis has been reported to cause 30-100% yield losses with estimated monetary loss of 52 million US dollars (Mahuku et al., 2015). Maize is a staple food crop and contributes to food security of farmers in sub-Saharan Africa. Given that this disease has spread quickly to a large geographical area in Sub-Saharan Africa within three years and no maize lines with complete resistance have been identified so far, this disease has a potential to cause significant losses in the future and needs to be studied and managed with proper care.

There are several other examples of crop epidemics that have impacted human civilization. Two notable examples which led to famines include the Irish potato famine of the 1840s and the Great Bengal famine of 1943 (Strange and Scott, 2005). The vast potential of pathogens to decrease crop yields and negatively impact society makes it important for geneticists and plant breeders to understand the genetic basis of resistance to crop diseases and continuously search for new sources of resistance. The remainder of this chapter provides an overview of the genetic architecture of major maize diseases and

a historical perspective about Goss's bacterial wilt and leaf blight, its re-emergence, and the genetic basis of resistance as it is currently understood.

1.2 History and re-emergence of Goss's wilt

Goss's wilt was first observed in Dawson county of south central Nebraska in 1969. It quickly spread to other areas of the state during the following years. By 1971, the disease was found in a seed corn field of western Iowa (Schuster et al., 1972b). Due to similar symptom morphology to Stewart's wilt, a bacterial disease of maize caused by *Erwinia stewartii*, Goss's wilt was thought to be caused by a highly virulent strain of *E. stewartii* during the early years. This newly discovered pathogen, however, produced discrete water soaking lesions with irregular margins -- referred to as 'freckles' -- instead of the continuous lesions caused by *E. stewartii* (Figure 1.1). The pathogen causing these leaf freckles was later identified and described as *Corynebacterium nebraskense,* which is now known as *Clavibacter michiganensis* subsp *nebraskensis* (*Cmn*) (Schuster et al., 1972b). This disease was originally referred to as "leaf freckles and wilt". However, in honor of prominent plant pathologist Dr. Robert Goss of the University of Nebraska, the disease was renamed as 'Goss's wilt and leaf blight'. Throughout this dissertation the abbreviated name 'Goss's wilt' will be used.

Figure 1.1 Water soaking lesions (i.e., "freckles) with irregular margins produced in response to inoculations with *Clavibacter michiganensis* subsp. *nebraskensis.*

From 1971 to 1974, this disease spread to the neighboring states including Iowa, Kansas, South Dakota, and Colorado (Schuster, 1975). The disease continued to spread in the next decade to other states throughout the Corn Belt and was reported in Nebraska, Iowa, Kansas, South Dakota, Colorado, Illinois, Minnesota, and Wisconsin by 1981 (Wysong et al., 1981). Maize germplasm was screened for resistance to Goss's wilt in several studies and varying levels of resistance was observed in inbred lines as well as hybrids, ranging from resistant to highly susceptible (Schuster et al., 1972a; Calub et al., 1974a; Wysong et al., 1981). Researchers and corn breeders quickly learned that varietal resistance was an effective measure to control the disease. Due in part to the deployment of partially resistant hybrids, Goss's wilt did not cause significant yield losses during late 1980s in commercial maize fields (Rocheford et al., 1989; Jackson et al., 2007a).

Rocheford et al., (1985) reported that many popcorn and sweet corn varieties were susceptible to Goss's wilt. It was speculated that selection for sugary-1 (*su1*) in sweet corn and gametophyte factor (*Ga*) in popcorn may have led to selection for a gene that caused susceptible reaction to Goss's wilt that was co-located with *su1* and *Ga* on

chromosome arm 4S. Wysong et al. (1981) found that only 3 out of 14 popcorn hybrids showed a high level of resistance and 10 out of 11 sweetcorn varieties showed severe Goss's wilt infestation. Further popcorn varieties were found to be highly susceptible in a recent Goss's wilt screening study conducted at the University of Nebraska (data not presented). For the next decade from 1990s to 2000s, Goss's wilt was reported in a sporadic manner from popcorn, sweetcorn, and susceptible dent corn fields (Jackson et al., 2007b).

Goss's wilt continued to spread within the mid western Corn Belt after it originated in 1969. However, it never reached its potential as pointed out by Wysong (1981) due to the ability of the plant pathologists and plant breeders to understand the host pathogen interaction and farmers being able to adopt management production practices. Goss's wilt however reemerged around 2006 as an important disease in the western Corn Belt and raised concerns for the maize seed industry.

In early 2006, the University of Nebraska Plant Disease Clinic at Scottsbluff, NE received maize samples from Nebraska, Wyoming, and Colorado which were diagnosed with Goss's wilt (Jackson et al., 2007b). Since then Goss's wilt has been spreading to other maize growing areas of North America. Goss's wilt has been officially confirmed in 13 states in the United States to date, including Nebraska, Iowa, Colorado, Missouri, Indiana, Illinois, Kansas, Minnesota, North Dakota, South Dakota, Wisconsin, Texas, and Louisiana (Ruhl et al., 2009; Malvick et al., 2010; Korus et al., 2011; EPPO, 2014; Friskop et al., 2014; Singh et al., 2015; Hosack et al., 2016) and Alberta, Ontario, and Manitoba provinces of Canada (EPPO, 2014; Howard et al., 2015). In 2013, Goss's wilt

was found in Louisiana (Singh et al., 2015) and Alberta (Howard et al., 2015), and in 2014 it was reported in Missouri (Hosack et al., 2016). The disease had never been reported in these states earlier, indicating that the range of Goss's wilt is expanding and could spread to other regions in the near future.

Several factors could have attributed to the re-emergence and spread of Goss's wilt (Jackson et al., 2007b; Harveson, 2012):

(i) Continuous corn cropping systems, which became more prevalent with higher corn prices.

(ii) Increased popularity of reduced tillage practices.

(iii) Increased frequency of susceptible hybrids, and reduced frequency of resistant hybrids sold by seed companies.

(iv) Conducive weather conditions.

(v) Increase in the use of center pivot irrigation systems, which could enhance pathogen survival and spread within the fields.

(vi) Emergence of new *Cmn* strains with increased virulence.

In a recent study, 40 environmental and agronomic factors that may have resulted in increased Goss's wilt incidence across the maize belt were tested for their contribution to Goss's wilt emergence (Langemeier et al., 2017). A multi-state survey was conducted to collect information on Goss's wilt incidence, environmental factors, and agronomic practices. The information collected about the above parameters was analyzed using classification, regression tree, and random forest analyses to identify the association of these factors with Goss's wilt development. Resistance of maize hybrids and planting

density were most closely associated with Goss's wilt incidence. Longitude, date of planting, crop rotation, surface residue, tillage, and growth stage were also predicted as important factors for emergence of the disease (Langemeier et al., 2016). Genetic and geographical diversity of *Cmn* may also be an important factor for Goss's wilt re-emergence and its expanded geographical range during the recent re-emergence. Using multi locus sequence typing (MLST) of 126 *Cmn* isolates, Webster, (2017) identified 23 sequence types in *Cmn* isolates and concluded that isolates collected from in the Northern Corn belt may have descended from the isolates from the Central Plains.

1.3 Causal organism of Goss's wilt: *Clavibacter michiganensis* subsp. *nebraskensis*

Clavibacter michiganensis subsp. *nebraskensis* is one of the eight subspecies of the only phytopathogenic species, *Clavibacter michiganensis* of the genus *Clavibacter* (Eichenlaub et al, 2007). Genus *Clavibacter* is a part of the Gram-positive actinomycetes and is a less studied genus as compared to Gram-negative bacteria (Eichenlaub et al, 2007).

The eight subspecies of *Clavibacter michiganensis* are very host specific and specifically infect only their hosts. These five subspecies include: 1) *Clavibacter michiganensis* subsp. *michiganensis* (*Cmm*) causes wilt and canker of tomato, 2) *Clavibacter michiganensis* subsp. *spedionicus* (*Cms*) produces ring rot of potato, 3) *Clavibacter michiganensis* subsp. *insidiosus* (*Cmi*) produces wilting and stunting of alfalfa, 4) *Clavibacter michiganensis* subsp. *tessellarius* (*Cmt*) causes leaf spots in wheat, and 5) *Clavibacter michiganensis* subsp. *nebraskensis* (*Cmn*) causes Goss's wilt and leaf blight in maize (Eichenlaub et al., 2007), 6) *Clavibacter phaseoli* causes bacterial bean

leaf yellowing, 7) *Clavibacter chiliensis*, and 8) *Clavibacter californiensis*. Although dent maize, popcorn and sweet corn are primary hosts of *Cmn*, it can infect other species such as *Sorghum bicolor* species (Shatter cane, Grain sorghum), *Sorghum x drummondii* (Sudangrass), *Saccharum officinarum* (Sugarcane), and *Alopecurus* (Foxtail) (Schuster, 1975; Langemeier, 2012).

The *Cmm* (tomato pathogen) and *Cms* (potato pathogen) were the first two *Clavibacter michiganensis* subspecies whose genomes were sequenced and annotated. The genome of one of the *Cmn* strains was later sequenced but not completely annotated. The genome size of the three subspecies were 3.30, 3.26, and 3.06 mega base pairs respectively. The GC content was found to be approximately 70% in all three subspecies. There were 106 insertion sequences in the genome of *Cms*, only three in *Cmm*, and no insertion elements were found in *Cmn*. The number of genes in the three subspecies was predicted to be equivalent with approximately 3,000 genes (Eichenlaub and Gartemann, 2011).

Earlier molecular studies to understand disease induction mechanism and host-pathogen interaction have been conducted with *Cmm*, and *Cms*. The genome of *Cmm* contains two circular plasmids (pCM1 and pCM2) and the genome of *Cms* contains a circular and a linear plasmid (pCS1 and pCSL1). The virulence factors *celA* and *pat-1* were identified on the pCM1 and pCM2 plasmids of *Cmm*. Eight homologues of *pat-1* on chromosome and three homologues on plasmids of *Cms* have been identified. Also, an orthologue of *celA* is present on plasmid pCS1 of *Cms* (Bentley et al., 2008). Along with these virulence factors on plasmids and chromosome, a pathogenicity island *chp*/*tomA*

has been identified on the chromosome of *Cmm* and is thought to be involved in disease induction. The *Cmn* strain NCPPB 2581, that has been sequenced, does not have any plasmids and no putative virulence factors have been identified in the genome (Eichenlaub and Gartemann, 2011). In an attempt to understand the genome organization of *Cmn*, the genomes of two other strains (DOAB 395 and DOAB 397) were sequenced and compared to the NCPPB 2581 strain. The genomes and proteomes of the three strains showed high level of similarity (Tambong et al., 2016). The mechanism of *Cmn* infection, therefore, still remains to be determined.

Clavibacter michiganensis subsp. *nebraskensis* is characterized as a gram positive, rod shaped bacterium which was first described as *Corynebacterium nebraskense* due to its coryneform or club shaped morphology (Vidaver and Mandel, 1974). The bacterial cells are non-motile and do not have flagella. *Cmn* grows slowly on nutrient media with visible appearance of apricot-orange colored colonies in three to four days. The bacteria can be cultured on nutrient-broth yeast extract agar, potato-dextrose agar, and synthetic media with yeast extract supplement. Among the three types of bacterial colonies observed by Vidaver and Mandel (1974), the most common colonies were circular, convex, shiny, and had a dark central spot differentiated from the outer margin. Unlike the most common colony type, the second most common colony type did not have an outer margin and the least frequent colony type was smaller than the others by a diameter of 2-4 mm. The optimum growth temperature for *Cmn* is 25-28 ^{0}C, with growth ceasing at 37 ^{0}C. *Cmn* is sensitive to 0.005% triphenyltetrazolium chloride agar

and susceptible to four bacteriophages. It has a GC content of 73.5% as reported from studies conducted on the NCPPB 2578 *Cmn* strain (Vidaver and Mandel, 1974).

Cmn is a residue borne bacterium and can survive in infested maize stubble for up to ten months when the stubble is left on the soil surface (Schuster, 1975). The pathogen enters into the plants primarily through wounds on the leaves and stems of plant typically caused by hail or wind damage (Jackson et al., 2007a). However, a recently observed *Cmn* infection was observed in Iowa maize fields, but no apparent wound or injury accompanied the infection, leading researchers to speculate that severe injury may not be required for *Cmn* to infect maize plants (Mallowa et al., 2015). A greenhouse experiment designed to follow up this observation indeed showed that epiphytic populations of *Cmn* on maize leaves can cause infection without severe wounding of the leaves. Scanning electron microscopy revealed that *Cmn* was associated primarily with the leaf areas that are sheltered, such as cuticle depressions, near the veins, epidermal cell junctions, cuticle cells near stomata, and the base of trichomes (Mallowa et al., 2015). While testing different plant parts of a susceptible line A632Ht for survival of *Cmn*, Biddle et al., (1990) detected *Cmn* in seeds along with stalks and ear shanks. Possible seed transmission of *Cmn* to the seedings was then tested by vacuum infiltrating the bacteria into seeds of maize and a very low transmission rate of 0.1 - 0.4 % to the seedlings from the seeds was observed (Biddle et al., 1990).

1.4. Symptoms and yield losses due to Goss's wilt

As the name 'Goss's bacterial wilt and leaf blight' indicates, Goss's wilt can cause two types of symptoms including: leaf blight and systemic wilt. The blight

symptoms appear as characteristic dark green to black water soaked discontinuous spots (freckles) on the leaves, which develop into lesions. As the disease progresses, lesions enlarge and bacterial exudates appear on the infected areas. The bacterial exudates dry to form a crystalline layer on the leaf surface that shines in sunlight. Coalescence of lesions eventually leads to blight of the leaves (Wysong et al., 1981) (Figure 1.2). If infection occurs early during the season, the bacterium can move systemically, eventually creating blockage in the vascular tissues and leading to wilt and possibly plant death if infection is severe enough. The bacteria can be seen as discoloration of vascular bundles examined from cross sections of the plant stems. Although wilt symptoms are more common if the plants are infected at the seedling stage, Goss's wilt can infect and can cause blight, wilt, and death of the plants at any stage (Wysong et al., 1981).

Figure 1.2 Leaf blight symptoms that developed after inoculations with *Cmn*.

Yield loss due to Goss's wilt can be severe if susceptible hybrids are planted. Over the last few years the Corn Disease Working Group (CDWG) consisting of plant pathologists across the maize growing states of the United States and Ontario province of

Canada has estimated the yield losses due to major diseases of maize. The estimates are based on statewide disease surveys, university extension feedback, and feedback from farmers. From 2012 to 2015, Goss's wilt has been ranked consistently among the top diseases in these surveys and yield losses of 38.5, 103.4, 215.9, and 139.8 million tonnes have been estimated to be lost to Goss's wilt respectively during these years (Mueller et al. 2016).

In earlier studies conducted during the 1970s and 1980s, yield losses due to Goss's wilt were recorded on both dent corn and sweet corn hybrids under artificial inoculations (Claflin et al., 1978; Pataky, 1988). Using artificial inoculation, a relatively resistant hybrid (B73 x Mo17) suffered only a 1% loss compared to its non-inoculated control, while a yield reduction of 44% was recorded on a susceptible hybrid (A619 x A632) as compared to its non-inoculated control (Claflin et al., 1978). Another study was conducted to assess the yield loss on a set of 42 closely related F_1 hybrids created using inbreds derived from A619 and A632. The minimum and maximum yield losses observed among the 42 hybrids were zero and 43.5 % respectively with an average yield loss of 18.6%. The loss of yield was significantly correlated to disease severity (Carson and Wicks, 1991). Yield reductions of 0 to 8.2% on resistant and greater than 17.4 to 40% on susceptible sweet corn hybrids were recorded over a three year period during 1984, 1985, and 1986 (Pataky et al. 1988).

1.5 Genetic architecture of maize disease resistance

Two major classes of genetic disease resistance are often distinguished in plants: 1) Qualitative resistance, also called major gene resistance or vertical resistance. In this

class, disease resistance is controlled by major resistance genes (R genes), which provide a complete resistance. 2) Quantitative resistance or horizontal resistance, which is controlled by a large number of small effect genes and is characterized by an intermediate level of disease resistance or a reduction in disease but not complete control or absence of disease (Parlevliet, 1981). The loci controlling quantitative disease resistance (QDR) are called as quantitative disease resistance loci (QRLs). It is believed that QDR is more durable than qualitative resistance because QDR is conferred by small effect of several genes and it would take a longer time for a pathogen to overcome multiple genes. This is because selection pressure is higher on the pathogen to mutate in the case of qualitative resistance (due to complete absence of disease) as compared to QDR (Parlevliet, 2002). While the biotrophic pathogens that feed on living host cells tend to be effectively controlled by qualitative resistance, necrotrophic pathogens that derive their nutrients from dead host tissues are often better controlled through QDR (Balint-Kurti and Johal, 2009). Possible explanation for rare occurrence of qualitative resistance to necrotrophs is that necrotrophic pathogens feed on dead tissue (Poland et al., 2009). The characteristic hypersensitive response (death of cells around point of infection) seen in cases of qualitative resistance can provide dead tissue for necrotrophs to feed on and make the plants susceptible.

Although the mechanism of QDR has not been completely understood, multiple possible mechanisms that may be involved in QDR were outlined by Poland et al., (2009). These six plausible hypotheses include:

(i) QDR is conditioned by pleiotropic effect of the genes that regulate plant morphology and development processes.

(ii) Pattern recognition receptors (PRR) such as receptor like kinases involved in basal defense may play a role in QDR.

(iii) Enzymes or phytoalexins released by plants to counteract toxins of plant pathogens may help in QDR.

(iv) QRLs may regulate signal transduction pathways such as salicylic acid, jasmonic acid, and ethylene that are involved in transduction of the defense signals.

(v) Quantitative resistance loci are attenuated forms of R genes.

(vi) Quantitative disease resistance is conditioned by genes that have not yet been identified.

According to Balint-Kurti and Johal (2009), unlike other major cereal crops including wheat and rice, only a handful of major genes for disease resistance have been discovered in maize and used by maize breeders. Examples of major resistance genes in maize cited by Balint-Kurti and Johal (2009) include *Ht* genes for resistance to northern corn leaf blight (Welz and Geiger, 2000), *Rp* genes for common rust resistance (Ramakrishna et al., 2002), *rhm1* gene for resistance to southern corn leaf blight (Chang and Peterson, 1995), and the *Rcg1* locus for anthracnose stalk rot resistance (Apraku, 1987). In addition to these major genes, maize breeders have relied on QDR for developing disease resistant cultivars. Wisser et al., (2006) published a review and meta-analysis of 50 publications reporting QTL for disease resistance in maize. A total of 437

QTL have been reported in these studies, out of which only 17 co-locate with resistance genes and 25 co-locate with resistance gene orthologs. The reported QTL span 89% of the maize genetic map and tend to cluster near chromosome ends.

Very recently, using the nested association mapping (NAM) population of maize, joint linkage mapping and genome-wide association studies have been conducted to help understand the genetic basis of resistance to major diseases of maize. The NAM population has an advantage over a single bi-parental population in that it can effectively capture the allelic variation in a diverse germplasm set. The maize NAM population consists of a common parent, B73, crossed to 25 diverse parents that capture a large amount of the genetic diversity present in maize (McMullen et al., 2009). Kump et al., (2011) evaluated 5000 NAM recombinant inbred lines (RILs) for resistance to SLB and reported 32 QTL with small additive effects that together explained 80% and 93% of the phenotypic and genetic variation respectively. Similarly, 29 QTL identified for northern corn leaf blight (NCLB) in the NAM population explained 77% of the phenotypic and 96% of the genetic variance. Estimated allelic effects for resistance at each of the QTL were small, and evidence for allelic series were found in which allelic effect sizes and directions were variable across different NAM families (Poland et al., 2011). Genetic control of resistance to gray leaf spot (GLS) was also studied using the NAM population. Sixteen QTL with small effect sizes were found, and interactions between three pairs of QTL were found to contribute to variation in GLS resistance (Benson et al., 2015).

Genome-wide association studies conducted using the NAM population provided high resolution to search for candidate genes within the genomic regions containing

significant SNPs. Several genes were identified that may play a role in plant disease resistance. Main classes of the candidate genes identified included: i) leucine rich repeat receptor-like kinases, which are known to be involved in disease resistance; ii) serine-threonine protein kinases, which play a role in plant basal defense; iii) mitochondrial carrier protein genes involved in programmed cell death; iv) genes with an antifreeze domain which are similar to pathogenesis related (PR) genes known for enhancing resistance; and v) AP2 transcription factors of the ethylene response factor (ERF) family, reported to regulate disease resistance pathways (Kump et al., 2011; Poland et al., 2011). Discovery of these genes for SLB, NCLB, and GLS have enhanced our understanding of QDR. The information about the number of genes controlling the Goss's wilt resistance is lacking as compared to the diseases discussed above. Hence, this dissertation will enhance our knowledge about genetic basis of Goss's wilt resistance.

1.6 Transcription profiling studies for diseases of maize

A transcriptome consists of the whole set of transcripts present in a cell or tissue. Major roles of transcriptomics, the study of the transcriptome of a species include: (i) characterization of different types of transcripts including mRNAs, non-coding RNAs, and small RNAs; (ii) determination of gene structure, alternate splicing patterns, and post transcriptional changes; (iii) quantification of changes in gene expression during development and in response to outside stimuli (Wang et al., 2009). RNA-seq has been used in maize previously in several studies to quantify gene expression levels in response to different stresses including drought (Kakumanu et al., 2012), nitrogen limitation (Bi et al., 2014), cold, heat, UV, and high salt (Makarevitch et al., 2015). However, only a few

studies have compared the response of resistant and susceptible lines of maize to pathogens using RNA-seq (Lanubile et al., 2014; Liu et al., 2016; WU et al., 2015). Lanubile et al. (2014) studied transcriptional changes in resistant CO441 and susceptible CO354 maize inbred lines at 72 hours after inoculation with *Fusarium verticillioides*. Resistant line CO441 showed higher level expression of genes belonging to different functional classes including pathogen perception, signaling and defense, and jasmonate and ethylene pathway. Higher expression of secondary metabolism pathway genes including shikimate, lignin, flavonoid, and terpenoid biosynthesis in CO441 was also observed, which indicated that selecting for these traits may improve *F. verticillioides* resistance.

Two QTL namely *qRfg1* and *qRfg2* have been discovered for resistance to *F. graminearum*, Gibrella stalk rot pathogen, which increased resistance by 32-43% and 12% respectively (Yang et al., 2010b). Liu et al. (2016) created three near isogenic lines (NILs) with *qRfg1*, *qRfg2*, and neither of the two QTL and conducted RNA-seq at 0, 6, and 18 hours post inoculation. Genes with functions in growth and development, photosynthesis, and defense were found to be differentially expressed. Similarly, WU et al. (2015) used resistant and susceptible NILs to identify the differentially expressed genes in response to *Pseudomonas syringae* pv. *syringae* (*Pss*) that causes brown spot disease of maize. Activation of large number of genes that are involved in PAMP triggered immunity, effector triggered immunity, and pathogenesis indicated that multiple defense pathways may be employed by maize for defense against *Pss* (WU et al. 2015).

1.7 Status of genetic studies on host plant resistance to Goss's wilt

Unlike other major maize diseases, genetic studies using molecular markers, genomics, and transcriptomics had not been applied to study Goss's wilt until recently because the disease appeared only sporadically and was not a major problem for commercial maize production. However, multiple historical studies had been conducted using classical genetics techniques. Gardner and Schuster (1974) evaluated 23 lines for Goss's wilt that were previously used in 13 x 13 and 10 x 10 diallel crosses. A resistant line and two susceptible lines used in one diallel and two resistant and two susceptible lines from the other diallel were selected for greenhouse testing for Goss's wilt resistance. The F_1, F_2, and backcross generations involving these 23 lines were also included in the experiment. A subsequent field experiment with the same plant materials was also conducted to determine if greenhouse testing is a good predictor of field testing. A correlation of 0.55 was found between greenhouse and field disease ratings, leading the authors to conclude that greenhouse ratings were not good predictors of field disease ratings. Overall, however, the lines found to be resistant in the greenhouse test were also found to be resistant in the field test. Progeny of the parents tested in these experiments were found to be intermediate to their parents in Goss's wilt resistance, and it was concluded that more than one major gene likely underlies resistance to this disease. The conclusion about the genetic basis of Goss's wilt was not definitive from this study, and further studies with improved techniques were recommended.

In order to better characterize the genetic architecture and predominant type of gene action controlling resistance to Goss's wilt, Martin et al. (1975) conducted a diallel

experiment including six parents (+1 in year 2) selected to be variable for resistance. A quantitative mode of inheritance was assumed and hence the diallels were analyzed with the Gardner and Eberhart (1966) model to determine the type of gene action involved. The sources of variation corresponding to the additive effects were determined to be significant in all cases (i.e., year combinations), while the source of variation corresponding to the dominance effects was significant in only one case (Martin et al., 1975). Breeding methodologies that focused on additive gene action and screening over several locations due to significant genotype-by-environment effects were recommended when breeding for resistance to Goss's wilt.

The mode of inheritance of Goss's wilt was further investigated with multiple diallels and generation mean analyses in both dent corn and sweet corn (Treat and Tracy, 1990; Treat et al., 1990; Ngong-Nassah et al., 1992). Treat and Tracy (1990), evaluated thirty-nine maize inbred lines for resistance to Goss's wilt, out of which 16 were classified as susceptible, 15 as intermediate, and eight as resistant. Two generation mean analyses (GMA), each consisting of a resistant and a susceptible line, and their F_1, F_2, and backcrosses to both parents indicated that additive gene action was the predominant type of gene action underlying resistance to Goss's wilt, accounting for 87 - 90 percent of the generation sums of squares. In two diallel experiments, one with five parents and one with six parents while GCA was highly significant, SCA was found to be non-significant. Similar results were obtained in diallel experiment of ten sweet corn hybrids from five sweet corn inbred lines in which GCA variance was much more important than SCA variance, as GCA and SCA accounted for 94% and 6% of the variation among crosses,

respectively (Treat and Tracy, 1990). GCA effects were once again found to be significant by Ngong-Nassah et al., (1992) in a diallel experiment conducted using the lines from the Northern Corn Belt, and explained 76% and 67% of the variation among crosses in greenhouse and field trials, respectively. All GMA and diallel studies indicated that additive gene action was more important for Goss's wilt resistance than dominance. Therefore, a recurrent selection approach was suggested as a strategy for increasing resistance to Goss's wilt in both dent corn and sweet corn germplasm (Treat and Tracy 1990).

Rocheford et al., (1989) used interchange (chromosomal reciprocal translocation) stocks to identify chromosomal arms that might carry genes for resistance. A series of interchange stocks in the background of resistant line M14 and susceptible line A632 were obtained and crossed to susceptible line A632 and resistant line Mo20W respectively. Additionally, F_2's, F_3's, and testcrosses were created. Progenies were evaluated for reaction to Goss's wilt in the greenhouse and field. The yellow endosperm marker used in the study was found to be associated with higher Goss's wilt symptoms in all experiments (field and greenhouse) involving F_2, F_3, and testcross progenies of M14 x A632 for an interchange on chromosome arm 4S. For A632 x Mo20W, a significant association was detected between an endosperm marker and disease rating in F_2 progeny for a chromosome 4S interchange. These results indicated that a gene controlling resistance to Goss's wilt is located on chromosome arm 4S. Besides chromosome arm 4S, significant associations of Goss's wilt symptoms with the yellow endosperm marker were also detected in some experiments for interchanges on chromosomes arms 8L and 7L.

The authors, therefore, did not rule out the presence of resistance genes on other chromosomes and concluded that results from this study support the previous results from diallel and generation mean analyses, which pointed towards the quantitative mode of inheritance of the disease. In an another attempt to estimate the number of genes Ngong-Nassah et al., (1992) estimated the number of genes involved in Goss's wilt resistance in F_2 populations. The number of estimated effective factors or genes in the six populations ranged from 0.79 to 4.60. The authors concluded that three to five genes that act in an additive manner may control the Goss's wilt reaction.

After a long gap of more than 20 years, the first reported use of molecular markers to study Goss's wilt resistance was published (Schaefer and Bernardo, 2013). In this genome-wide association study (GWAS), a collection of 284 diverse maize inbred lines adapted to Minnesota was used to find quantitative trait loci (QTL) for different traits including flowering time, kernel composition, NCLB, and Goss's wilt. Nine SNPs were found to be significantly associated with Goss's wilt diseased leaf area in this GWAS. Chromosomes 1, 4, 5, and 9 each contained two significant SNPs and one SNP was located on chromosome 7. Nine SNPs together explained 47% of the phenotypic variation for Goss's wilt. The allelic effects of major alleles at each SNP was small and ranged from -1.12 to 0.68 on a square root transformed scale of 0 - 100 percent.

1.8 Genomic selection for disease resistance breeding

Quantitative disease resistance as discussed in section 1.5 is typically controlled by several small effect genes. Breeding strategy for QDR would be different than breeding for qualitative disease resistance (Poland and Rutkoski, 2016). Two approaches

that are commonly used for breeding qualitative disease resistance are (i) forward breeding or population improvement by increasing the frequency of desirable alleles in the population, (ii) backcross breeding to introgress the resistant gene or a big effect QTL from resistant parent to the susceptible parent. The backcross strategy is most useful when a resistance gene is present in a wild accession and it need to be transferred to an elite variety that lacks the gene. Given that QDR is conferred by several small effect genes, use of backcross breeding approach can be ruled out and forward breeding or population improvement through recurrent selection would be better approaches for breeding QDR into crop plants.

Population improvement for QDR is gradual and requires multiple cycles of selection (St.Clair, 2010). Different methodologies of forward breeding that can be used for incorporating QDR include (i) phenotypic screening of a large number of individuals in early generations to discard the susceptible progeny, (ii) marker assisted selection that uses molecular markers linked to the resistance genes to increase the frequency of desirable alleles in the population (Dekkers and Hospital, 2002), (iii) Genomic selection that uses genome-wide molecular markers and can increase genetic gains per unit time (Meuwissen et al., 2001).

While phenotypic selection has worked for plant breeders to develop new varieties without completely understanding the genetic control of the traits, it takes a long time (at least 10 years) to release a cultivar (Tanksley et al., 1989) and setting up of disease screening nurseries can be laborious. Marker assisted selection (MAS) is faster than phenotypic selection alone and can be used to increase genetic gain per unit time in

recurrent selection programs by increasing the frequency of favorable alleles in a breeding population at QTL of interest. Marker assisted recurrent selection (MARS) can accelerate the breeding process with the use of continuous off season nurseries and availability of genotypic information at the seedling stage as three to four cycles of MARS can be completed in one year (Eathington et al., 2007). However, MARS is a two step approach that involves quantitative trait loci (QTL) mapping to identify tightly linked molecular markers to the disease trait which are then used in selection. It requires validation of QTL and is difficult when a trait is controlled by a large number of small effect QTL. Small QTL effects are inconsistent and difficult to detect and pyramiding QTL into a desired line is difficult as the number of QTL increases (Bernardo, 2010). Improvement in the phenotype is minimal unless many of the small effect QTL are successfully pyramided into a single line.

Genomic selection (GS) uses molecular markers uniformly distributed throughout the genome to build a prediction model using both phenotypic data and genotypic data from the training population. The predictive statistical model is then used to predict the genotypic values of the individuals of the validation population, which would be genotyped only and for which no phenotypic data is available (Meuwissen et al., 2001). Unlike linkage mapping and genome-wide association mapping, genomic prediction does not involve a QTL detection step but rather estimates marker effects for all markers simultaneously and uses all marker effects in the calculation of predicted genetic values (Bernardo and Yu, 2007). If genomic prediction models can accurately predict genetic values, rates of genetic gains could be increased through genomic selection by reducing

the breeding cycle time through the application of molecular marker information ahead of phenotyping (Poland and Rutkoski, 2016). Using eight bi-parental maize populations Beyene et al., (2015) showed that cycle 3 hybrids developed through GS yielded 7.3% higher than the hybrids developed from conventional pedigree selection.

Instead of using significant markers only as is the case in MARS, whole genome markers in GS context can be used to improve the frequency of desirable alleles in breeding populations. This approach was called rapid cycle genomic selection (RCGS) and was implemented by the International Maize and Wheat Improvement Center's (CIMMYT) global maize breeding program to increase the efficiency of recurrent selection programs (Babu et al., 2012). CIMMYT uses RCGS to improve multi-parent populations, which are created by crossing elite maize inbred lines in all possible combinations. Each inbred line carries a specific trait such as drought tolerance, disease resistance, and nutritional quality. Half diallels are made from 8 - 12 maize inbred lines and the resulting F_1's are intermated in isolation blocks to obtain a large F_2 population. In total, 500 S_2 families (C0, cycle 0 population) are created for each multiparent population and are testcrossed to a tester. Testcrosses are phenotyped in subsequent season and the top 10% of the C0 families are selected and then recombined to form C1 (cycle 1). Individuals of C1 are genotyped and their genomic estimated breeding values are estimated using marker effects from C0 (Babu et al., 2012).

Initial studies of genomic prediction in maize have been focused on traits such as grain yield, grain dry matter, biomass, grain moisture, and plant height (Albrecht et al., 2011; Riedelsheimer et al., 2012). However, genomic prediction has been applied to

several diseases of maize as well as other crops. For example, *Fusarium* head blight (FHB) in barley (Lorenz et al., 2012), FHB in wheat (Rutkoski et al., 2012), stem rust of wheat (Rutkoski et al., 2015), sudden death syndrome of soybeans (Bao et al., 2015), and cassava mosaic virus (Wolfe et al., 2016). In maize, NCLB (Technow et al., 2013), Maize lethal necrosis (MLN) (Gowda et al., 2015), *Fusarium* ear rot (Zila, 2014) and *Gibberella* ear rot (GER) (Riedelsheimer et al., 2013) have been predicted with genomic prediction.

Genomic prediction accuracy was assessed for diseases of maize both in diversity panels as well as breeding populations. A maximum prediction accuracy of 0.71 was obtained by Technow et al. (2013) for NCLB in a breeding population and minimum prediction accuracy of 0.36 was observed for maize lethal necrosis in a diversity panel (Gowda et al. 2015). Several factors have been evaluated in GS studies for increasing the prediction accuracy. For example, Technow et al. (2013) concluded that NCLB resistance can be improved through genomic selection and combining of individuals from the heterotic groups can increase prediction accuracy significantly. An improvement in prediction accuracy for MLN and *Fusarium* ear rot was observed when significant SNPs from GWAS were included as fixed effects in the GS model (Zila, 2014; Gowda et al., 2015). In a dataset of five interconnected bi-parental doubled haploid populations, prediction accuracies of severity of *Giberella* ear rot and deoxynivalenol (DON) increased with increasing training population size and decreased when full sibs were replaced by half sibs (Riedelsheimer et al., 2013). The percent of variation explained by genomic predictions from realized genomic relationship matrices was significantly higher than that explained by each of the significant SNPs from GWAS for MLN and *Fusarium*

ear rot (Zila, 2014; Gowda et al., 2015). This showed that while selecting for specific significant SNPs may not provide large gains in resistance levels due to small effect of each SNP, using whole genome markers can significantly improve the breeding populations for resistance. Chapter four of this dissertation will explore the prospects of GS for Goss's wilt resistance breeding.

In summary, although the genetic basis of resistance to Goss's wilt has been studied using classical quantitative genetics methodologies such as diallel and generation mean analysis, molecular markers and QTL mapping have not been used to study the genetic architecture of Goss's wilt. The main objective of this dissertation is to understand the genetic basis of resistance to Goss's wilt using quantitative genetics and genomics methodologies including linkage mapping and joint linkage mapping, genome-wide association mapping, and RNA-sequencing. As no major QTL was detected in linkage mapping and genome-wide association mapping, prospects of genomic prediction as a strategy for Goss's wilt resistance breeding have been explored in chapter four of the dissertation.

Chapter two of this dissertation has been published in Crop Science.

Author contributions:

Designed the experiment: AJL. Conducted field experiments: AS, APA. Analyzed the data: AS. Wrote the manuscript: AS. Assisted in inoculum preparation: TAJZ. Edited and reviewed the manuscript: AJL.

Chapter Two: Mapping Quantitative Trait Loci for Resistance to Goss's Bacterial Wilt and Leaf Blight in North American Maize by Joint Linkage Analysis

Amritpal Singh, Aaron P. Andersen, Tamra A. Jackson-Ziems, and Aaron J. Lorenz*

A. Singh, A.P. Andersen, and A.J. Lorenz, Dep. of Agronomy and Horticulture, Univ. of Nebraska, Lincoln, NE 68503-0915; T.A. Jackson-Ziems, Dep. of Plant Pathology, Univ. of Nebraska, Lincoln, NE 68503-0722. Received 8 Sept. 2015. Accepted 21 Mar. 2016.

*Corresponding author (alorenz2@unl.edu).

2.1 Abstract

Goss's wilt and leaf blight is a bacterial disease of maize (*Zea mays* L.) caused by the Gram-positive bacterium *Clavibacter michiganensis* subsp. *nebraskensis*. Goss's wilt has re-emerged as an important disease in the western United States and is spreading to other areas. Although the reasons for this re-emergence are not completely known, it is important to understand the genetic basis of resistance to Goss's wilt. The objective of this study was to map the quantitative trait loci (QTL) underlying resistance to Goss's wilt. To achieve this objective, joint linkage and linkage mapping in 3 of the 25 nested association mapping families were used. Three biparental linkage mapping families including 'B73' × 'Oh43', B73 × 'HP301', and B73 × 'P39' were evaluated for Goss's

wilt in Nebraska. Eleven QTL were detected on chromosomes 1, 2, 3, 4, 5, and 10 through joint linkage mapping. The joint linkage model explained 45% of the phenotypic variation for Goss's wilt. Linkage mapping in each of the three families identified nine, six, and four QTL in the families B73 × Oh43, B73 × HP301, and B73 × P39, respectively. Joint linkage and linkage analysis were also conducted within each environment to detect any environment-specific QTL. However, most of the QTL were colocalized with QTL detected in across-environment joint linkage and linkage mapping. These results will help us to understand the genetic basis of resistance to Goss's wilt better and may facilitate maize breeding programs to incorporate resistance to Goss's wilt into the maize germplasm.

Abbreviations: NAM, nested association mapping; QTL, quantitative trait loci; RIL, recombinant inbred line; SNP, single nucleotide polymorphism; WMD, weighted mean disease.

2.2 Introduction

Clavibacter michiganensis subsp *nebraskensis* is a Gram-positive bacterium causing Goss's wilt and leaf blight disease in maize (Vidaver and Mandel, 1974). Goss's wilt was first discovered in south central Nebraska in 1969 and quickly spread to different counties within Nebraska, as well as to Iowa, Kansas, Colorado, and South Dakota. In the disease evaluation trials, varying levels of resistance were found in maize inbred lines (Calub et al., 1974; Schuster et al., 1972). Partially resistant hybrids were developed and through the deployment of resistant hybrids over a period of 10 yr, the disease became sporadic, did not cause any severe damage, and occurred only in fields

planted with susceptible maize hybrids (Jackson et al., 2007a; Vidaver et al., 1981). However, around 2006, Goss's wilt re-emerged as an important disease in the North American Corn Belt (Jackson et al., 2007a). Since then, Goss's wilt has been observed in 60 counties in Nebraska (Jackson and Rees, 2010) and 80 counties in Iowa (Robertson, 2012). Goss's wilt has also been reported in Colorado, Illinois, Iowa, Kansas, Michigan, Minnesota, South Dakota, and Wisconsin (CABI and EPPO, 2000).

Goss's wilt infection can occur at any developmental stage, with the bacteria entering into the plant through leaves, stems, and roots (Schuster, 1975). There are two possible phases of Goss's wilt: (i) leaf blight and (ii) systemic wilt. Infection typically starts through wounds on the plant surface leading to water-soaked spots (or freckles), which progress to gray lesions and eventually lead to leaf blight. Less common systemic wilt occurs when the pathogen infects the vascular system and moves systemically through the xylem, leading to blockage of the vascular bundles. The systemic wilt phase is more common if infection occurs during early growth stages (Jackson et al., 2007a).

Loss in yield can be severe if Goss's wilt occurs on susceptible hybrids. Yield loss caused by Goss's wilt under artificial inoculation was estimated to be 44% in the susceptible maize hybrid 'A619' × 'A632' but the tolerant hybrid B73 × 'Mo17' yielded only 1% less than the control (Claflin et al., 1978). Similarly, yield loss was estimated to be 43.5% in a study including 42 related hybrids derived from the inbred lines A632 and A619. Yield reduction was found to be correlated with disease severity (Carson and Wicks, 1991). A significant reduction in yield was found in a susceptible sweet corn hybrid when inoculated at early as well as later growth stages, though yields of resistant

hybrids were not affected. Disease incidence and severity depended on the level of resistance of the hybrid (Suparyono and Pataky, 1989a). Although recent yield losses caused by Goss's wilt are not well documented, it has been estimated that yield losses from Goss's wilt in Iowa, Illinois, Minnesota, and Nebraska totaled 0.878 Tg (Mueller and Wise, 2012).

Inheritance of resistance to Goss's wilt has been studied previously using classical techniques. Although screening maize inbred lines and their F_1, F_2, and backcross generations (Gardner and Schuster, 1974) concluded that no maize inbred line was completely resistant to Goss's wilt, crosses were intermediate in performance and more than one gene lay behind variation in resistance. In a subsequent study, Goss's wilt resistance again appeared to be a polygenic trait (Martin et al., 1975). Later studies using diallel mating designs and generation mean analysis indicated that additive variation accounted for most genotypic variation for Goss's wilt (Ngong-Nassah et al., 1992; Treat and Tracy, 1990).

Identification of strong associations between markers and resistance would be highly desirable, as phenotyping for Goss's wilt is laborious and prone to failure because of a variety of possible weather conditions, including hot and dry weather around inoculation. Moreover, a marker–QTL analysis would be the starting point for ultimately discovering the genes controlling resistance. To identify the chromosomal arms that possibly harbor the genes conferring resistance to Goss's wilt, (Rocheford et al., 1989) screened 'M14' interchange stocks and found strong evidence of a genetic factor on chromosome arm 4S but were not able to rule out the presence of other resistance genes

on other chromosome arms. Recently, in a genome-wide association study for different traits in historical Minnesota maize inbred lines, nine single nucleotide polymorphisms (SNPs) were found to be significantly associated with resistance to Goss's wilt that together explained 47% of the phenotypic variation (Schaefer and Bernardo, 2013).

Joint analysis of multiple QTL mapping populations with a common parent increases the power and precision of QTL detection and can better capture allelic variation in a comparatively diverse set of germplasm compared to single biparental mapping populations (Blanc et al., 2006). One such population is the nested association mapping (NAM) population of maize, in which 25 diverse founder lines were crossed with a common inbred line, B73 (Yu et al., 2008). Using the maize NAM population, numerous QTL with small additive effects have been identified for diseases of maize such as southern corn leaf blight and northern corn leaf blight (Kump et al., 2011; Negeri et al., 2011; Poland et al., 2011). These studies have greatly enhanced knowledge of the genetic architecture underlying these traits. Such studies are lacking for Goss's wilt, mainly because of this disease's sporadic occurrence for many years. The recent re-emergence of Goss's wilt, however, has increased interest in developing a better understanding of the genetic basis of resistance.

The objective of this study was to use joint linkage mapping to map the QTL underlying resistance to Goss's wilt in three diverse maize genetic backgrounds. The distribution of QTL effects among three types of maize was examined. The results from this study will provide knowledge on the genetic architecture underlying variation for

resistance to Goss's wilt and will ultimately help to inform marker-based selection strategies and searches for resistance loci.

2.3 Material and Methods

2.3.1 Germplasm

Seed of F_5–derived recombinant inbred lines (RILs) from three families; B73 × Oh43, B73 × HP301, and B73 × P39 was obtained from the maize genetics stock center and increased through sib mating during the summers of 2011 and 2012 in Lincoln, NE. These families of RILs were developed as a part of the NAM project (Yu et al., 2008). Each family consisted of 200 RILs. The common parent, B73, is moderately resistant to Goss's wilt but the other three parents are believed to be relatively susceptible. A second reason why these parents were chosen because that the inbred lines Oh43, HP301, and P39 were derived from dent corn, popcorn, and sweetcorn backgrounds, respectively. The RILs have previously been genotyped with 1106 SNP markers (http://panzea.org, accessed 28 Apr. 2016).

2.3.2 Field experiment

In 2012, 195 RILs from the B73 × Oh43 family were planted in a completely randomized design with one replication at O'Neill, NE. The RILs were planted in single-row plots, 3.7 m long and 0.8 m apart. Inbred lines B73, Oh43, and A632 were planted as replicated checks to estimate the experimental error and make spatial adjustments as needed. Each check was replicated seven times.

In 2013, 172 RILs from the B73 × Oh43 family, 141 RILs from the B73 × HP301 family, and 125 RILs from the B73 × P39 family were planted at Mead, NE. Most RILs

were planted in one replication in a completely randomized design, except for 54 RILs from the B73 × Oh43 family, 53 RILs from the B73 × HP301 family, and 63 RILs from the B73 × P39 family. These RILs were replicated twice because extra field space and seeds were available. The inbred lines B73, Oh43, HP301, P39, and B14A were replicated six times each as checks throughout the experiment.

In 2014, 174 RILs from the B73 × Oh43 family, 143 RILs from the B73 × HP301 family, and 124 RILs from the B73 × P39 family were planted in a completely randomized design. As there was more interest in the dent types, the B73 × Oh43 family was replicated twice in 2014, but the other two families were planted in only one replication. The inbred line B14A (susceptible) was planted as a replicated check 23 times throughout the experiment to compare the success of inoculations and make spatial adjustments if necessary. The number of RILs across the years varied according to seed availability. In summary, the B73 × Oh43 family was planted and evaluated in 2012 (one replicate), 2013 (partially replicated), and 2014 (two replications). The B73 × P39 and B73 × HP301 families were evaluated in 2013 (partially replicated) and 2014 (one replicate).

2.3.3 Inoculation and disease rating

Clavibacter michiganensis subsp. *nebraskensis* isolates were maintained on nutrient broth agar media. The isolates were tested on the susceptible sweetcorn variety ‘Golden Cross Bantam’ in the greenhouse for pathogenicity before using them for inoculum preparation. Inoculum for field inoculations was prepared from five *Clavibacter michiganensis* subsp. *nebraskensis* isolates consisting of approximately 3 ×

10^8 colony-forming units per mL. These isolates were collected as part of a multistate survey across the Midwest (Langemeier, 2012). Only isolates collected in Nebraska were used for field inoculations (225A, 225B, 225C, 10B, and 194C) according to Animal and Plant Health Inspection Service regulations. In 2012, inoculations for the B73 × Oh43 family at O'Neill, NE were performed by DuPont Pioneer using proprietary techniques. One disease rating was recorded at O'Neill 54 d after inoculation using a disease rating scale of 1–9 (Suparyono and Pataky, 1989b) on a whole-plot basis (Supplementary Fig. 2.1), where 1 represents a symptomless plot, 2 indicates disease spread within 5 cm of the point of inoculation, 3 indicates limited disease spread over 5 cm from the point of inoculation, 4 indicates a large spread and lesions extending to other end of the inoculated leaf, 5 indicates systemic infection with blight of uninoculated leaves; 6 indicates leaf blight and wilt, 7 and 8 indicate severe leaf blight and wilt, and 9 represents a completely dead plot. In 2013, inoculations were performed 49 d after planting when most of the plants were at the V6 stage of development. Plants were wounded using motorized weed whippers (curved shaft string trimmer, Model UT33600, Homelite Consumer Products Inc., Anderson, SC). Inoculum was sprayed on the plants within 10 s after wounding using a backpack sprayer. Disease ratings were recorded 15 d and 30 d after inoculations using the rating scale described earlier. In 2014, inoculations were performed using the same method used in 2013 but the ratings were recorded 15, 30, and 45 d after inoculations.

2.3.4 Phenotypic data analysis

Weighted mean disease (WMD), which is equivalent to the standardized area under the disease progress curve (Balint-Kurti et al., 2010), was calculated for each environment separately. This was calculated as the average of two consecutive ratings multiplied by the number of days between the ratings. The values were then summed and finally divided by the total number of days of evaluation to obtain the WMD (Balint-Kurti et al., 2010). When only two ratings were taken, the mean of the two ratings was equal to WMD. Analysis of variance (ANOVA) was performed on the WMD values using PROC MIXED (SAS version 9.3, SAS Institute Inc., Cary, NC). Assessment of the check performance in different areas of each field indicated that systematic spatial variation was not present and thus no spatial adjustments were applied. Least-squares means were calculated by fitting a model including environment and RIL as fixed effects. Pearson's correlation coefficients among environments were calculated using PROC CORR (SAS version 9.3). Heritability on a per-plot basis for Goss's wilt was estimated for each of the three families separately using PROC MIXED (SAS version 9.3) according to the method given by (Holland et al., 2003).

2.3.5 Linkage and joint linkage mapping

Quantitative trait locus mapping was conducted using the WMD values for each plot and the SNP data on the three linkage mapping families. Joint stepwise regression, implemented in GLMSELECT (SAS version 9.3), was used to build a model of cofactors, where environment, family, and marker nested within family effects were fit as fixed effects (Buckler et al., 2009). The level of significance for effects to enter and remain in

the model was set to p-values of 0.0001 and 0.0002, respectively. Cofactor selection was also conducted separately for each environment.

The entire genome was scanned using a window size of 20 cM with the cofactors identified in the model described above. One thousand permutations were used to determine the logarithm of odds threshold to maintain an experiment-wise error rate of 0.05 (Doerge and Churchill, 1996). The logarithm of odds threshold was determined to be 4.18. After identifying significant markers, significant allelic effects were tested for significance at $P < 0.05$ using a *t*-test comparing the alternative parent allele to the founder parent (B73) allele. To calculate the variation explained by each QTL, a general linear multiple regression model was fitted with environment, family, and significant marker effects using PROC GLM (SAS version 9.3).

In addition, QTL mapping was performed on each linkage family separately across environments and within each environment. Instead of joint stepwise regression, as implemented earlier for all three families combined, a stepwise regression model was fitted with environment effects and marker effects only for each family using GLMSELECT (SAS version 9.3; Buckler et al., 2009). A one-dimensional scan of the genome was conducted in the same way as described above for the joint analysis. Linkage and joint linkage analyses were implemented using a SAS script previously available at the Buckler Lab website (www.maizegenetics.net, accessed August 2013).

2.4 Results and Discussion

2.4.1 Phenotypic distribution

The phenotypic distribution of Goss's wilt was skewed towards resistance because of a lack of systemic disease development in most of the RILs (Fig. 2.1). This is likely to point towards the difficulty of establishing good artificial infection and the fact that the V6 (later) growth stage was targeted in the inoculation method used in this study. Infection at earlier growth stages through wounding of plants through hail and winds has been reported to cause severe Goss's wilt symptoms and yield losses (Jackson et al., 2007b). The inbred line B73 was found to be resistant compared to Oh43 and HP301. However, B73 and P39 did not differ significantly for Goss's wilt ratings. A genotype × environment interaction was found to be highly significant in the combined dataset ($P < 0.0001$) and for the B73 × Oh43 family ($P < 0.0001$) (Table 2.1). Variation caused by a genotype × environment interaction was not significant for the B73 × HP301 and B73 × P39 families. Earlier studies have reported significant genotype × environment interactions for Goss's wilt and have advised several years of testing when selecting genotypes for resistance to Goss's wilt (Carson and Wicks, 1991; Ngong-Nassah et al., 1992; Treat et al., 1990). Significant positive correlations were found among years for Goss's wilt WMD in all families. Correlations among the three environments for Goss's wilt ratings for B73 × Oh43 family ranged from $r = 0.63$ to 0.71 ($P < 0.0001$). Correlations for Goss's wilt ratings between the two environments were also significant for the B73 × HP301 family ($r = 0.60$, $P < 0.0001$) and the B73 × P39 family ($r = 0.61$, $P < 0.0001$). High positive correlations indicated consistency in disease development across

environments despite the skewed distribution of phenotypes and hence the data were combined across environments for the QTL analysis as well as being analyzed separately. Carson and Wicks (1991) also observed high correlations between Goss's wilt ratings recorded in different years despite the presence of a hybrid × environment interaction. Heritability estimates on an individual plot basis for Goss's wilt were high and very similar across the three families, ranging only from 0.60 to 0.62. In a single-year trial of F_2 populations, broad-sense heritability estimates for Goss's wilt were also high (0.63–0.80) in resistant × susceptible crosses. Heritabilities have been reported to be lower in intermediate × susceptible crosses (0.21–0.33) (Ngong-Nassah et al., 1992).

2.4.2 Linkage and joint linkage mapping

This study is the first to report QTL for Goss's wilt resistance using linkage mapping techniques. Linkage and joint linkage mapping across environments detected 11 QTL controlling resistance to Goss's wilt (Fig. 2.2; Table 2.2). The allelic effect estimates were small, especially in joint linkage mapping, where no allelic effect was greater than 0.5 on a rating scale of 1 to 9 (Fig. 2.3). Previous studies using diallel, generation means analysis and chromosomal interchange stocks indicated that the inheritance of resistance to Goss's wilt is polygenic (Ngong-Nassah et al., 1992; Rocheford et al., 1989; Treat and Tracy, 1990). The results from this study, as well as those reported by Schaefer and Bernardo (2013), are in accordance with this hypothesis, as each QTL identified explained only a small amount of phenotypic variation, ranging from 1 to 6% (Table 2.2).

Single-family linkage mapping in the B73 × Oh43 family identified nine QTL on chromosomes 1, 2, 3, 4, 9, and 10. In the B73 × HP301 family, six QTL were detected on chromosomes 1, 2, 5, 6, 7, and 10. In the B73 × P39 family, four QTL were detected on chromosomes 2, 4, 5, and 9 (Fig. 2.2). All QTL detected via linkage mapping in the B73 × Oh43 family were also detected with joint linkage mapping. However, linkage mapping in the B73 × HP301 family detected QTL on chromosomes 6 and 7 that were not found using joint linkage mapping. Similarly, four QTL on chromosomes 1, 2, 4, and 9 were detected using single-family linkage mapping in the B73 x P38 family, but were not detected using joint linkage mapping (Fig. 2.2). The statistical significance of these QTL just exceeded the thresholds in the single-family analysis, and their statistical significance in the joint linkage analysis was just below the threshold. This can probably be attributed to the lack of an effect at these positions within the B73 × Oh43 family, which was the largest family and was evaluated in the most environments and thus contributed the most data. The absence of an effect in the B73 x Oh43 family could have diluted the effect within the smaller families, resulting in a lack of significance in these loci of relatively small effect. However, due to the unbalanced nature of the data in this study, the exact explanation is unknown.

All but two of the QTL were detected using the cross-environment analysis, which was expected on the basis of the high correlations between environments for Goss's wilt ratings. Two additional QTL were detected when the analyses were performed on single environment, but these QTL were small (R^2 = 0.04 and 0.07) and were only detected using single-family analyses (Supplementary Table 2.1). This result

suggests these QTL are stable across environments and that perhaps QTL for Goss's wilt resistance detected in general show little interaction with the environment.

The motivation for including sweetcorn and popcorn parents was to find alleles that make sweet corn and popcorn susceptible to Goss's wilt compared to B73 but we found that all parents contributed alleles conferring both resistance and susceptibility (Fig. 2.3, Table 2.2, Supplementary Table 2.1). Allelic effect estimates of the QTL were positive at some loci but negative at others, indicating that B73, although relatively resistant, carries alleles for susceptibility to Goss's wilt (Fig. 2.3, Table 2.2). For example, the allelic effect of B73 is negative (susceptible) on chromosome 1 at 94.9 cM and positive (resistant) on chromosome 4 at 111.3 cM (Fig. 2.3). This was expected, as transgressive segregation was observed in each family. On chromosome 1 at 134 cM, the B73 allele had a positive effect in the B73 × Oh43 family but a negative effect in the B73 × HP301 family (Fig. 2.3). This indicates the presence of an allelic series and possibly a "common gene–rare variant" situation as observed for other traits in the maize NAM (Wallace et al., 2014).

In conclusion, we report several QTL associated with resistance to Goss's wilt and their allelic effects across three distinct genetic backgrounds. Both joint linkage and linkage mapping helped in identification of the QTL. The QTL may be useful to maize breeders attempting to introgress resistance to Goss's wilt into elite lines used in dent corn, popcorn, and sweetcorn breeding.

2.5 Acknowledgements

We are grateful to Dow AgroSciences for providing the funding and making this study possible. We thank DuPont Pioneer for carrying inoculations at their O'Neill location in 2012. The authors are also thankful to the Lorenz lab members: Nonoy Bandillo, Dnyaneshwar C. Kadam, Jon Luetchens, M.D. Liakat Ali, and Ben Deaver for their help in Goss's wilt phenotyping and the Jackson lab members: Jae Brungardt, Brad Tharnish, and Cody Kaarstad for their help in inoculum preparation.

Table 2.1 ANOVA of Goss's wilt infection data of all three families of maize (B73 × Oh43, B73 × HP301, and B73 × P39) combined, as well as independently.

Source of variation	df	Mean Square	*F*-value	*P*-value
		Combined dataset		
Environment	2	45.07	139.29	<0.0001
Family	2	4.77	14.75	<0.0001
RIL(family)	448	2.95	9.12	<0.0001
RIL(family) × Environment	530	0.70	2.15	<0.0001
Residual	279	0.32	–	–
		B73 × Oh43		
Environment	2	27.96	87.31	<0.0001
RIL	188	5.00	15.63	<0.0001
RIL × environment	325	0.88	2.75	<0.0001
Residual	185	0.32	–	–
		B73 × HP301		
Environment	1	35.51	143.75	<0.0001
RIL	140	1.33	5.40	<0.0001
RIL × environment	113	0.36	1.46	0.0802
Residual	42	0.25	–	–
		B73 × P39		
Environment	1	4.35	10.96	0.0017
RIL	120	1.61	4.06	<0.0001
RIL × environment	90	0.38	0.96	0.5682
Residual	52	0.40	–	–

Table 2.2 Significant genetic markers from joint linkage mapping and linkage mapping in each family of maize conducted across environments.

Marker†	Chr‡	Pos§ (cM)	LOD¶	2-LOD# (cM)	R^2††	Additive effect‡‡		
						Oh43	HP301	P39
Joint linkage mapping								
an1.5	1	94.9	29.5	92.0–96.5	0.06	0.433	0.060	0.238
PHM4942.12	1	134	5.3	120.3–137.6	0.01	0.166	0.157	0.030
PZA02957.5	1	176.9	11.4	164.6–177.5	0.02	0.222	0.160	0.216
PZA00902.1	2	6.8	16.6	6.4–8.5	0.03	0.328	0.083	0.008
PZA03559.1	2	41.8	18.7	41.5–47.7	0.04	0.368	0.094	0.018
PZA02017.1	2	106.2	14.1	105.3–107.5	0.03	0.273	0.088	0.206
PZA00494.2	3	97.8	13.0	92.6–101.4	0.03	0.384	0.108	0.018
PZA02479.1	4	111.3	20.0	110.4–112.2	0.05	0.306	0.302	0.101
PZB01017.1	5	74.5	15.6	71.3–75.6	0.03	0.144	0.349	0.295
PZB00547.3	9	40	7.8	34.5–46.6	0.02	0.201	0.050	0.125
PZA03196.1	10	48.8	18.3	43.8–53.2	0.04	0.257	0.330	0.099
Linkage mapping (B73 × Oh43)								
an1.5	1	94.9	30.8	92.0–108.4	0.10		0.502	
PZA02204.1	1	171.4	7.6	164.6–177.5	0.03		0.254	
PZA00902.1	2	6.8	14.0	5.7–10.0	0.05		0.330	
PZA03559.1	2	41.8	18.7	41.5–49.8	0.06		0.412	
PHM3668.12	2	106.1	9.8	103.7–115.3	0.04		0.279	
PHM824.17	3	100.5	10.4	84.6–103.0	0.04		0.462	
PZA02479.1	4	111.3	14.2	108.7–112.5	0.06		0.279	
PZB00547.3	9	40	6.2	34.5–45.2	0.02		0.345	
PZA00647.9	10	52.2	6.1	43.8–56.1	0.02		0.209	
Linkage mapping (B73 × HP301)								
PHM4531.46	1	39.7	7.9	37.8–43.2	0.07		0.268	
PZA03747.1	2	22.6	5.4	11.5–27.6	0.05		0.226	
PZA00934.2	5	56.1	5.4	45.4–66.8	0.05		0.237	
PHM15961.13	6	0	4.0	0.0–10.7	0.03		0.189	
PZA02274.1	7	135	5.7	121.1–135.0	0.05		0.226	
PZA01005.1	10	49.2	6.1	47.1–63.0	0.05		0.247	
Linkage mapping (B73 × P39)								
PZA03577.1	2	154.9	5.1	142.7–155.7	0.06		0.263	
PZA03203.2	4	57.4	5.0	55.4–60.6	0.06		0.289	
PZA01779.1	5	68.1	14.3	66.8–72.5	0.19		0.494	
PZA00466.1	9	20.7	4.8	12.6–21.0	0.06		0.249	

† Marker name as listed on nested association mapping map in cM

‡ Chromosome

§ Map position of each marker on the chromosome

¶ Logarithm of odds score

Two-logarithm of odds (2-LOD) support interval in cM

†† Variation explained by each marker

‡‡ Additive effect estimates of the alleles from each parent

Supplementary Table 2.1 Significant genetic markers from joint linkage mapping and linkage mapping in each family conducted for each environment/year separately. The columns from left to right display marker name, chromosome, map position of each maker on the chromosomes in centimorgan (cM), logarithm of odds (LOD), and 2-LOD support interval in cM, variation explained by each term (R2), and additive effect estimate of alleles.

[a]Marker	[b]Chr	[c]Pos (cM)	[d]LOD	[e]2-LOD (cM)	[f]R^2	[g]Additive effect		
						Oh43	HP301	P39
Joint Linkage mapping 2013								
PZA02393.2	1	33.1	7.1	31.7-37.8	0.03	-0.192	0.244	0.041
PZA02823.1	1	133.9	7.4	129.8-134.8	0.03	-0.251	0.095	0.094
PZD00022.5	2	155.7	7.6	152.5-155.7	0.04	-0.048	-0.160	0.316
PZA03647.1	3	96.9	15.9	92.6-103.2	0.08	-0.453	0.080	-0.157
PZA01926.1	4	69.8	7.3	61.8-76.2	0.03	0.200	-0.125	-0.428
PZA00155.1	4	111.5	8.3	102.9-112.2	0.04	0.213	0.306	0.109
PZA00067.10	5	72.5	21.9	68.7-74.5	0.11	-0.159	-0.353	-0.596
PZA00758.1	8	49.9	5.8	42.0-52.4	0.03	0.258	-0.018	0.232
PZA00466.1	9	20.7	6.6	18.7-28.5	0.03	0.314	0.079	0.130
PZA02398.2	10	43.4	6.8	40.6-53.2	0.03	0.257	0.233	0.023
Joint linkage mapping 2014								
an1.5	1	94.9	17.9	92.0-108.4	0.12	-0.445	-0.078	-0.255
PZA00894.7	1	180.9	6.3	169.2-188.2	0.04	0.236	0.157	0.200
PZA03559.1	2	41.8	5.4	38.6-62.2	0.03	0.271	-0.108	0.058
PZB00772.7	2	117.5	4.2	109.9-125.9	0.03	0.186	-0.162	0.177
PZB01017.1	5	74.5	7.0	72.5-75.6	0.04	-0.233	-0.228	-0.223
PZA03196.1	10	48.8	7.7	44.8-53.2	0.05	0.168	0.387	0.152
Linkage mapping B73 x Oh43 2012								
PZA00455.14	1	96.5	9.6	89.6-108.4	0.15		-0.716	
PZA00497.4	2	49.8	5.1	41.5-64.2	0.07		0.513	
PHM3637.14	4	92.7	6.2	81.9-102.6	0.09		0.560	
PZA02128.3	10	44.8	7.4	42.9-53.2	0.11		0.642	
Linkage mapping B73 x Oh43 2013								
PZA03228.4	2	50.8	4.1	41.5-58.8	0.06		0.290	
PHM824.17	3	100.5	9.5	96.7-103.2	0.15		-0.456	
PZA03275.4/1	4	85.2	5.4	81.7-89.1	0.08		0.333	
PZA00416.7	8	20.7	4.7	10.5-32.1	0.07		0.309	
Linkage mapping B73 x Oh43 2014								
an1.5	1	94.9	16.7	92.0-98.4	0.18		-0.478	
PZA00978.1	1	177.5	3.6	164.6-191.5	0.03		0.210	
PZA01211.1	2	10.0	4.2	0-22.6	0.04		-0.232	
PZA03559.1	2	41.8	8.0	41.5-49.8	0.08		0.352	
PZA00494.2	3	97.8	5.9	92.6-101.4	0.06		-0.273	
PZA03645.1	7	49.4	3.4	47.8-63.7	0.03		0.208	
Linkage mapping B73 x HP301 2013								
PHM4531.46	1	39.7	6.5	29.9-43.2	0.10		0.311	

PZA01935.10	2	19.5	3.5	7.1-27.6	0.05	-0.232
PZA00155.1	4	111.5	7.1	110.4-112.2	0.11	0.341
PZA01796.1	5	75.6	7.9	71.3-78.4	0.13	-0.359
PZA00048.1	10	42.9	6.8	38.6-46.7	0.11	0.322
Linkage mapping B73 x HP301 2014						
PZA03274.4	5	50.8	3.6	45.4-66.8	0.10	-0.30808
PZA03713.1	10	48.0	5.4	44.8-58.4	0.16	0.370403
Linkage mapping B73 x P39 2013						
PZA01735.1	2	91.5	4.2	85.5-105.3	0.08	0.310686
PZA03203.2	4	57.4	5.1	55.4-75.3	0.09	-0.35472
PZA01779.1	5	68.1	9.3	66.8-74.5	0.19	-0.47276

a Marker name as listed on NAM map in centimorgam

b Chromosome

c Map position of each marker on the chromosome

d Logarithm of odds score

e 2-LOD support interval in centimorgan

f Variation explained by each marker

g Additive effect estimates of alleles from each parent

Figure 2.1 Histograms showing distributions of the least-squares means of Goss's wilt recombinant inbred lines of maize. Both combined and individual family distributions are displayed.

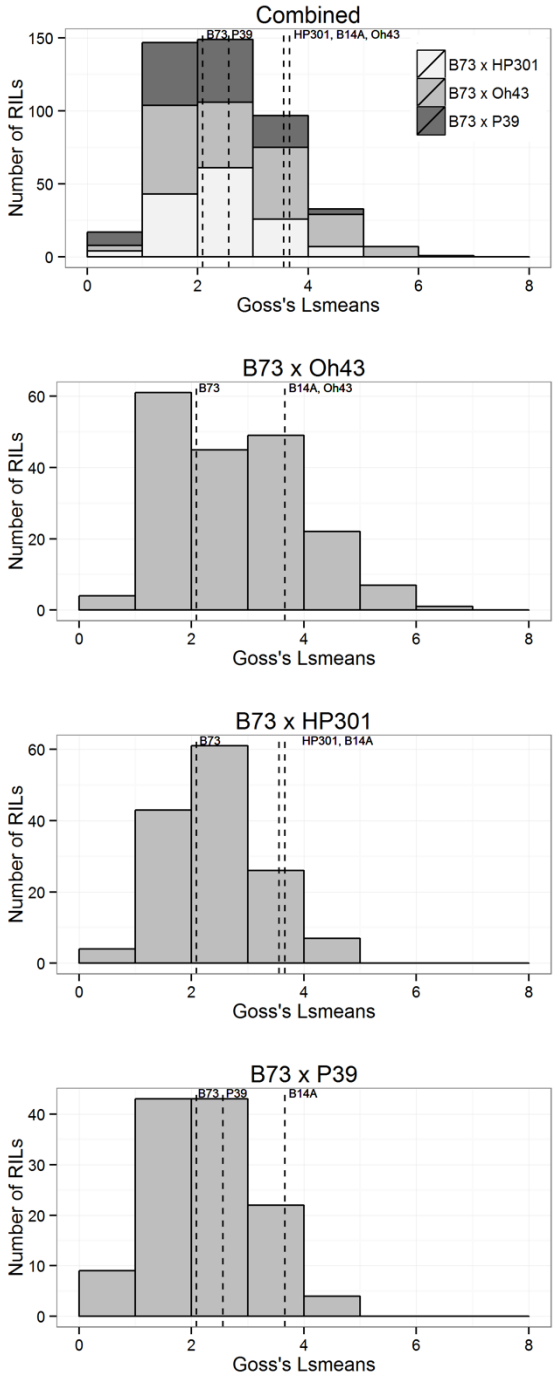


Figure 2.2 Comparison of quantitative trait locus (QTL) positions from joint linkage mapping and linkage mapping in each family. Ten maize chromosomes are shown as vertical gray bars. Segments of different colors indicate mapped QTLs at that position either identified using joint linkage or linkage mapping. The lengths of the segments show the range of the two logarithm of odds (2-LOD) support interval of the QTLs.

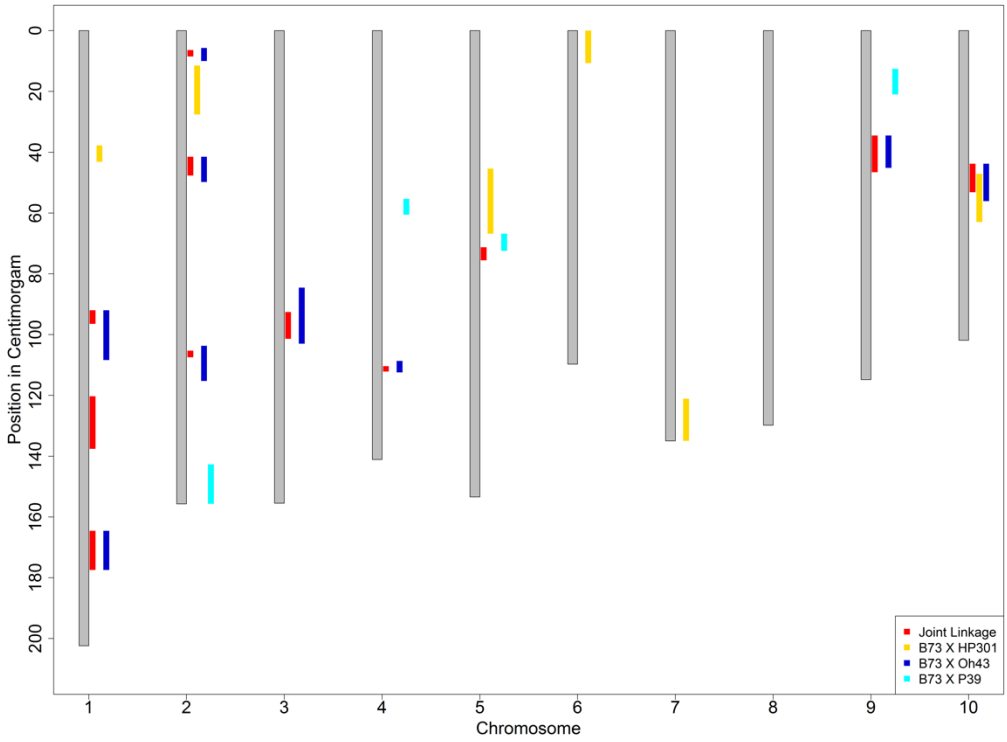
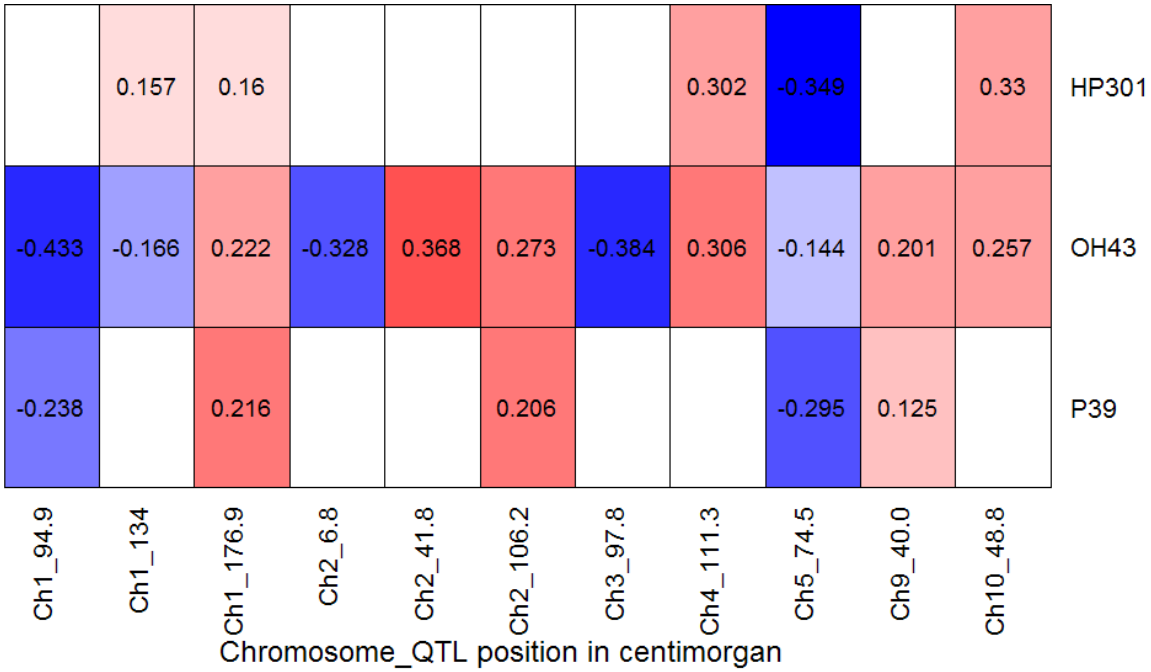


Figure 2.3 Heat map of allelic effect estimates for the B73 allele and founder maize parents from joint linkage mapping across environments. Positive (red) effects indicate that the B73 allele confers increased resistance and negative (blue) effects indicate that the B73 allele contributes to susceptibility. Only allelic effects that are significantly different from zero at the 5% significance threshold level are colored.



Supplementary Figure 2.1 Picture of maize leaves displaying variable Goss's wilt symptoms to illustrate the rating scale used for this study. Ratings 1 through 8 are shown. Only completely dead plants were given a rating of 9.

Chapter Three: A Genome-Wide Association and Gene Co-Expression Network Analyses Reveals Complex Nature of Resistance to Goss's Wilt of Maize

AUTHORS: Amritpal Singh, Alex Brohammer, Guangyong Li, Candice Hirsch, James Alfano and Aaron J. Lorenz

Amritpal Singh, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108.

Alex Brohammer, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108.

Guangyong Li, Department of Plant Pathology, University of Nebraska, Lincoln, NE 68503-0722.

Candice Hirsch, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108.

James Alfano, Department of Plant Pathology, University of Nebraska, Lincoln, NE 68503-0722.

Aaron J. Lorenz, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108

3.1 Abstract

Goss's wilt and leaf blight is a bacterial disease of maize caused by the gram positive bacterium *Clavibacter michiganensis* subsp. *nebraskensis* (*Cmn*). First discovered in Nebraska, Goss's wilt has now spread to major maize growing states in the United States and three provinces in Canada. Previous studies conducted using elite maize inbred lines and their hybrids have shown that resistance to Goss's wilt is a quantitative, polygenic trait. Candidate genes linked to resistance have yet to be identified. Discovery of resistant alleles and incorporation into maize hybrids can effectively control the disease. The objective of this study was to gain a better understanding of the genetic basis of resistance to Goss's wilt of maize by using a combined approach of genome-wide association mapping and gene co-expression network analysis. Genome-wide association analysis was conducted using both a diversity panel consisting of 555 maize inbred lines, and a set of 450 recombinant inbred lines (RILs) from three bi-parental mapping populations connected by a common parent. Genotype data consisted of 342,237 single nucleotide polymorphisms (SNPs) scored using genotyping-by-sequencing (GBS). Three SNPs in the diversity panel and 10 SNPs in the combined dataset of diversity panel and RILs were found to be significantly associated with Goss's wilt resistance. Each significant SNP explained 1% to 5% of the phenotypic variation for Goss's wilt. Potential candidate genes underlying the associations were identified. To augment the results of genome-wide association mapping and further identify the genes that may be involved in response of maize to *Cmn*, a time course RNA sequencing experiment was conducted using resistant (N551) and

susceptible (B14A) maize inbred lines. Gene co-expression network analysis of this time course experiment identified two modules of correlated genes that showed differential regulation in response to *Cmn* inoculations between N551 and B14A. Gene ontology analysis revealed that one of the modules was enriched in defense-related functions. Results from this study significantly improved our understanding of the genetic basis of Goss's wilt and identified SNPs that can be validated for use by maize breeding programs to impart resistance to Goss's wilt.

3.2 Introduction

Goss's bacterial wilt and leaf blight of maize, caused by a gram positive bacterium *Clavibacter michiganensis* subsp. *nebraskensis* (*Cmn*) has recently re-emerged in the United States Mid-Western Corn Belt (Jackson et al., 2007b). Since its re-emergence around 2006, Goss's wilt has continued to spread throughout North American maize growing regions. In the United States Goss's wilt has been reported in Nebraska, Iowa, Colorado, Missouri, Indiana, Illinois, Kansas, Minnesota, North Dakota, South Dakota, Wisconsin, Texas, and Louisiana, and in Canada it has been confirmed in Alberta, Manitoba, and Ontario (Ruhl et al., 2009; Malvick et al., 2010; Korus et al., 2011; EPPO, 2014; Friskop et al., 2014a; Howard et al., 2015; Singh et al., 2015; Hosack et al., 2016). Infection by *Cmn* can dramatically reduce grain yield. Artificial inoculations of *Cmn* conducted during the 1990s showed yield losses up to 44% (Carson and Wicks, 1991), and yield losses of 50% or more were reported during recent epidemics of Goss's wilt (Robertson, 2012). In recent years Goss's wilt ranked among major diseases of maize in North America in terms of yield loss estimates, with estimates ranging from

38.5 million bushels in 2012 to 215.9 million bushels in 2014 in major maize producing U.S. states and Canada (Mueller and Wise, 2012, 2014).

Identifying sources of resistance to Goss's wilt and deploying resistance into maize hybrids is the best strategy to reduce yield losses (Treat et al., 1990). A great degree of variation in level of resistance to Goss's wilt exists in maize. Maize inbred lines have shown variable response to Goss's wilt varying from resistant, intermediate to highly susceptible based on screenings in the 1970s and 1980s with a limited number of inbred lines (Schuster et al., 1972a; Calub et al., 1974a; Wysong et al., 1981). B14 and its derived lines, such as A619, have been found to be generally susceptible; Oh43 was reported to be moderately susceptible; and Mo17 was reported to be resistant (Schuster et al., 1972a; Calub et al., 1974a). Classical quantitative genetics mating designs and analyses, including diallels and generation means analyses, indicated that resistance to Goss's wilt is under polygenic control with the genetic variation being primarily additive genetic variation (Gardner and Schuster, 1974; Martin et al., 1975; Rocheford et al., 1989; Treat and Tracy, 1990; Ngong-Nassah et al., 1992).

Efforts to identify molecular markers linked to resistance to help to unravel the genetic architecture of Goss's wilt have been lacking until recently because occurrence of this disease was only sporadic during the development and adoption of molecular markers and QTL mapping by plant geneticists and breeders. The re-emergence of Goss's wilt has increased interest in identifying sources of resistance, molecular markers that could be used for selection, and genes controlling resistance. We previously conducted linkage mapping using three bi-parental maize populations and identified eleven QTL of

small effect, half of which were population specific (Singh et al., 2016). While linkage mapping with bi-parental populations has high power for QTL detection, this technique can only resolve QTL to 10-20 cM, making fine mapping necessary to better resolve the QTL (Holland, 2007). On the other hand, association mapping provides better resolution, can be helpful in identification of causal genes for a trait of interest, eliminates the time required to develop a mapping population, and has an ability to simultaneously survey all the allelic diversity contained in the germplasm of interest (Yu and Buckler, 2006).

Both candidate gene-based association mapping and genome-wide association mapping have proven to be powerful methods for discovery of candidate genes underlying quantitative disease resistance (QDR) in maize (Kump et al., 2011; Wisser et al., 2011; Zila et al., 2013). For example, a glutathione S-transferase (*GST*) gene was found to be associated with resistance to three maize diseases -- northern corn leaf blight, grey leaf spot and southern corn leaf blight – using a multivariate association mapping approach (Wisser et al., 2011). Similarly, Zila *et al.* (2013) discovered potential candidate genes associated with *Fusarium* ear rot of maize using a genome-wide association analysis on a maize diversity panel. However, no candidate genes have been found to be associated with resistance to Goss's wilt. Schaefer and Bernardo (2013) performed genome-wide association mapping for Goss's wilt and identified several SNPs associated with resistance, each explaining no more than 10% of the phenotypic variation. The diversity panel used by Schaefer and Bernardo (2013), however, was fairly limited in size (n = 284) and diversity, being comprised of only elite maize inbred lines bred by public universities (mostly the University of Minnesota) or private seed companies. This limited

diversity, combined with relatively few molecular markers, lowered the resolution of mapping. Search for candidate genes or QTL in the regions of significant SNPs revealed that a significant SNP in chromosome bin 1.10 was co-localized with a QTL for Stewart's wilt. Nevertheless, no other candidate genes could be associated to Goss's wilt resistance (Schaefer and Bernardo 2013).

Gene expression analysis have been used extensively to study the genetic architecture and mechanism of disease resistance in plants (Matsumura et al., 2003). Many hybridization and sequence based approaches have been developed to study gene expression (Wang et al. 2009). RNA-sequencing is a useful approach that can be used to quantify the changes in gene expression during development and response to stimuli with several advantages over traditional methods (Wang et al. 2009). Combination of differentially expressed genes in response to biotic and abiotic stresses and results from QTL mapping and genome-wide association studies (GWAS) can possibly give us a broader, more comprehensive view of the genetics underlying quantitative disease resistance. In maize, Lanubile et al. (2014) used RNA-seq to identify transcriptional changes in response to *Fusarium verticillioides* in resistant CO441 and susceptible cultivars CO354. Higher levels of expression of genes with gene ontology functional categories such as pathogen perception, secondary metabolism, and signaling and defense were observed in CO441.

Weighted gene co-expression network analysis (WGCNA) is a systems biology approach that can be used to analyze RNA-seq data to find co-expression patterns among the genes (Langfelder and Horvath, 2008). Instead of focusing on individual genes,

WGCNA can be used to identify co-expressed sets of genes (known as modules) in response to any treatment. For example, genes inside a particular module can show an increased or decreased level of expression in response to a pathogen. Functional enrichment analysis of modules of genes can be done with Gene Ontology software to identify if the genes in a particular module are enriched in resistance related functions.

The objectives of this study were: 1) Characterize the genetic variation for Goss's wilt within a large panel of diverse inbred lines; 2) Discover genomic regions and candidate genes controlling variation for resistance to Goss's wilt; 3) Compare the variation explained by detected loci in the GWAS to the variation explained by a genomic relationship matrix in order to better describe the genetic architecture of Goss's wilt resistance; 4) Discover modules of genes that show changes in gene expression patterns between a resistant and susceptible maize line in response to inoculation with *Cmn*. The differentially regulated gene modules were compared to results from the GWAS to provide additional evidence of candidate genes discovered by the GWAS, as well as quantify the amount of variation in resistance explained by differently expressed gene modules, providing additional valuable information on the genetic architecture underlying this complex trait.

3.3 Methods and Materials

3.3.1 Germplasm and selection of diversity panel for genome-wide association analysis

Two sets of germplasm were used for GWAS mapping: a diversity panel of 555 inbred lines and three bi-parental linkage mapping populations selected from the maize

nested association mapping (NAM) population. The diversity panel of 555 maize inbred lines was selected from a larger set of 2815 maize inbred lines genotyped by Romay *et al.* (2013). This panel consisted of lines from Stiff Stalk Synthetic (SSS), non-SSS, tropical, popcorn, and sweet corn genetic backgrounds. A priori subpopulation classification (i.e., SSS, non-SSS, popcorn) of each line was taken from Romay et al. (2013). The first year of Goss's resistance evaluation included 400 inbred lines selected from the set of 2815. To arrive at the set of 400, the set of 2815 was first reduced to 900 lines by retaining only the lines that reached silking within four growing degree days of B73 according to silking growing degree data provided by Romay et al. (2013). This was done in order to reduce variation in days to flowering and hence reduce any confounding effects variation in days to flowering could have on disease ratings. A *k*-means clustering analysis was then applied to SNP data to classify the 900 lines into 400 clusters and one line from each of the 400 clusters was randomly selected. Seed of the 400 inbred lines was obtained from the North Central Regional Plant Introduction Station (NCRPIS) in Ames, IA and increased during the summer of 2014 by self pollination. Low power to detect associations in the first year of this study prompted an increase in the panel size for evaluations in 2015. The panel was increased to 555 by adding 155 inbred lines from the set of 900 selected on silking date criteria described above and for which adequate seed was immediately available.

The three NAM bi-parental populations selected were B73 x Oh43, B73 x HP301, and B73 x P39. These same populations were used in an earlier study to identify QTL for Goss's wilt through linkage mapping (Singh et al., 2016). While the hub parent, B73, is

moderately resistant, other three parents were found to be comparatively susceptible to Goss's wilt in a preliminary screening. Moreover, Oh43, HP301, and P39 are dent, popcorn, and sweetcorn types, respectively, which allowed the study of Goss's wilt resistance in three distinct genetic backgrounds. In 2012, 195 RILs from B73 x Oh43 family were screened for Goss's wilt. In 2013, 172 RILs from B73 x Oh43 family, 141 RILs from B73 x HP301 family, and 125 RILs from B73 x P39 family were evaluated. In 2014, 174 RILs from B73 x Oh43 family, 143 RILs from B73 x HP301 family, and 124 RILs from B73 x P39 family were phenotyped for Goss's wilt. The lines were evaluated using completely randomized field design with replicated checks as described earlier (Singh et al. 2016).

3.3.2 Genotypic data

A dataset consisting of 681,257 GBS SNPs downloaded from panzea.org in 2013 was used to select the diversity panel. In 2014, ZeaGBSv2.7 version dataset of 955,690 GBS SNPs was made available at panzea.org, which was then used for the genome-wide association analysis described below. In brief, the GBS data was obtained by Romay et al. (2013) using the *ApeKI* restriction enzyme as previously described (Elshire et al., 2011). A reference genome-based GBS pipeline in TASSEL software for SNP discovery with standard parameters as applied to maize Discovery Build was used to call the SNPs (Glaubitz et al., 2014). The imputed version of ZeaGBSv2.7 dataset was used which was imputed using Fast Inbred Line Library InputatioN (FILLIN) method. For imputing the missing data, FILLIN generates high coverage haplotypes from inbred lines and then uses

these haplotypes to impute the target samples by identifying the genetically closest haplotype donor to the sample being imputed (Swarts et al., 2014).

3.3.3 Diversity panel characterization

Population stratification within the diversity panel was visually assessed using principal component analysis (PCA) and ADMIXTURE, a software for model-based estimation of ancestry among the individuals (Alexander et al., 2009). A cross validation procedure implemented in ADMIXTURE was used to initially choose the optimum number of subpopulations (K) by minimizing the cross validation error. Multiple runs of ADMIXTURE were conducted at different values of K ranging from 3 to 20. Twenty replications were performed for each value of K. Cluster memberships for each replicate were aligned using the cluster matching software CLUMPP (Jakobsson and Rosenberg, 2007). A plot of cross validation error versus K was examined and it was determined that the optimum value of K was between three and five, but an exact optimum could not be determined. K was set to three based on visual inspection of a PCA plot of PC1 vs PC2, subpopulation membership plots from ADMIXTURE differing in K, and subpopulation information from Romay *et al.* (2013). The average cluster membership across twenty replications was used as a covariate in the GWAS model.

Decay of linkage disequilibrium (LD) was assessed as pairwise correlation between the SNP markers on each chromosome using PLINK (Purcell et al., 2007). Pairwise correlations were calculated between SNP pairs within a 10 kilo-base (kb) window. To assess the relationship among the lines of the diversity panel and visualize the clustering of inbred lines according to Goss's wilt resistance, a distance matrix was

created with PLINK in which the distance was expressed as genomic proportions i.e. 1-identity-by-state (1-IBS) and a neighbor joining tree was created using the *ape* package in R (Paradis et al., 2004). Subpopulations within the diversity panel were color coded in the neighbor joining tree. In order to visualize the patterns of resistance and susceptibility to Goss's wilt by subpopulations, inbred lines were color coded on the neighbor joining tree based on their disease rating. A spectrum of colors was assigned to the inbred lines labels with green indicating resistance and red indicating susceptibility.

3.3.4 Goss's wilt phenotyping and disease nursery

In 2014, the diversity panel of 400 inbred lines was planted in a Goss's wilt nursery at the Agricultural Research and Development Center of the University of Nebraska in Mead, NE. Plots were arranged in a randomized complete block design with three replications. Susceptible line B14A as well as two susceptible and two resistant proprietary check inbred lines from Dow AgroSciences were included to assess disease development. In 2015, the diversity panel of 555 lines was planted at the same location using the same experimental design as in 2014. Inoculations with *Cmn* were carried out following the same procedure as described previously (Singh et al., 2016). Briefly, wounds were created on plant leaves with motorized weed whippers and *Cmn* inoculum was sprayed within seconds of injuring the plants to ensure infection. Disease ratings were recorded 15, 30, and 45 days after inoculations (DAI). A disease rating scale of 1 to 9 used by Singh et al. (2016) on a whole plot basis was used in this study, where 1 represents complete resistance, 2 indicates disease spread less than approximately 5 cm from the point of inoculation, 3 represents limited spread but more than 5 cm from the point of

inoculation, 4 indicates a large spread with lesions often extending middle of the leaf, 5 indicates systemic infection and lesions on un-inoculated leaves, 6 indicates blight of un-inoculated leaves and wilting of plants, 7 indicates severe blight and wilt, 8 indicate severe blight and severe wilt with a limited green tissue on leaves and stems of plants, and 9 represents completely dead plot.

3.3.5 Phenotypic analysis

Three visual ratings taken after inoculation at 15, 30, and 45 DAI were combined to calculate weighted mean disease (WMD) scores. For calculation of WMD, the average of two consecutive ratings was taken and multiplied by the number of days between the two ratings. These values were summed and divided by the total number of days spanning the first and last rating (Balint-Kurti et al., 2010; Singh et al., 2016). Analysis of variance (ANOVA) was conducted on WMD values using ASReml-R (Butler et al., 2009) by fitting the following model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + r_{k(i)} + \gamma_{ij} + \varepsilon_{ijk} \quad [1]$$

Where, Y_{ijk} represents the WMD value, μ is the grand mean, α_i is the effect of year *i*, β_j is the effect for inbred line *j*, $r_{k(i)}$ is the effect of the k^{th} replication nested within the i^{th} year, γ_{ij} is the interaction effect between inbred line and year, and ε_{ijk} is the residual. All effects besides the residual were treated as fixed effects. Best linear unbiased estimates (BLUEs) for WMD of the inbred lines were calculated. For estimating heritability, inbred line, inbred line-by-year interaction, and residual variances were estimated using ASReml-R by fitting these effects as IID random effects. Plot-based heritability was

calculated as $H^2 = \sigma_G^2 \Big/ (\sigma_G^2 + \sigma_{GY}^2 + \sigma_\varepsilon^2)$ where σ_G^2 is the variance among inbred lines, σ_{GY}^2 inbred-by-year interaction variance, and σ_ε^2 is the residual variance.

3.3.6 Genome-wide association model

Genome-wide association mapping was performed using the following mixed linear model:

$$y = X\beta + Wm + Zu + e \quad [2]$$

where *y* is a vector of WMD BLUEs of the inbred lines; β is a vector of fixed subpopulation (K = 3) effects; *X* is an incidence matrix relating β to *y* and contains subpopulation membership probabilities output from ADMIXTURE; *m* is a vector of fixed SNP effects; *W* is a marker matrix indicating the allelic state of each inbred line for each of the marker effects included in *m*; *u* is a vector of random polygenic effects where $u \sim MVN(0, G\sigma_u^2)$ and *G* is a kinship matrix calculated using the marker data; *Z* is a design matrix relating *u* to *y*; and *e* is a vector of random residuals where $e \sim MVN(0, I\sigma_e^2)$. Model [2] was implemented using the factored spectrally transformed linear mixed model (FaST-LMM) algorithm (Lippert et al., 2011). Kinship matrix was calculated from all the markers used in GWAS by adding -sim parameter to FaST-LMM command. The linear mixed model described was also applied to the combined dataset of diversity panel and bi-parental populations. The model was modified to include a fixed environmental effect to account for the different environments in which these germplasm sets were evaluated. Also, subpopulation effects were extended to include subpopulation effects for each of the three bi-parental populations. The subpopulation effect incidence matrix, *X*, which included subpopulation membership probabilities as described above,

was extended by adding three columns for each of the bi-parental populations, where each column contained a 1 when the RIL in that row belonged to the corresponding bi-parental population and a 0 elsewhere.

In order to declare SNPs as significantly associated with Goss's wilt, a false discovery rate (FDR) based on a q-value of 0.1 was used (Storey and Tibshirani, 2003). To calculate the percent variation explained (R^2) by significant SNPs after accounting for subpopulation effects, a multiple regression model was fit including Goss's wilt BLUEs as the dependent variable, and subpopulation effects and effects of significant SNPs as independent variables. The total variation in Goss's wilt resistance for the diversity panel that could be explained by the kinship matrix (*G*), also referred to as "genomic heritability", was calculated using the GBLUP model in the rr-BLUP R package by fitting a kinship matrix created from all SNP markers as random effect. The variation explained by kinship matrix was compared to the variation explained by the significant SNPs only.

3.3.7 Haplotype analysis

Once the significant SNPs were identified by GWAS, a haplotype analysis was performed in the regions of the genome surrounding the significant SNPs using the software Haploview (Barrett et al., 2005). The SNPs within 10 kb of each significant SNP were included initially to conduct the haplotype block analysis. Haplotype blocks were defined according to the four gamete rule which is based on the idea that if all four gametes are observed for a pair of SNPs, a recombination has taken place between the adjacent SNPs assuming absence of backward mutation (Wang et al., 2002). A cutoff of

1% was used to define the haplotype block boundaries, meaning that if a fourth two-SNP haplotype allele was observed at a frequency of greater than 1%, recombination was assumed to have occurred between the SNPs that formed the haplotype. Allele frequencies of haplotype alleles were examined within each subpopulation of the diversity panel to determine the allele frequency differences among subpopulations.

3.3.8 Plant materials and inoculations for transcriptome profiling

Transcriptome profiles were obtained for B74A and N551. B14A was found to be susceptible to and N551 was found to be resistant *Cmn* in the field screening described above. Average rating of B14A and N551 were 3.4 and 1.0 respectively in the field on 1(resistant) to 9 (susceptible) rating scale. B14A and N551 both belong to the stiff stalk heterotic group and are closely related (B14 was a founder of the synthetic population from which N551 was derived) (Russell, 2006).

To perform inoculations in general, seeds of the inbred lines were planted in plastic inserts and kept in the greenhouse for two weeks before inoculations. Inoculations were performed when the plants were at V2 stage. For inoculations, plants were transported to the lab and were inoculated using a vacuum infiltration method with the Welch 1400 Duo Seal Vacuum Pump. Inoculum was prepared from *Cmn* isolate 12038, which was tested and determined to be virulent on maize. The bacterial cells were suspended in 10mM $MgCl_2$ for measuring the concentration and then were mixed into distilled water. The inoculum bacteria concentration was set to 1 x 10^8 colony forming units with a spectrophotometer. In order to increase the surface tension of the suspension, 0.005% Tween20 was added to the inoculum. The plants were placed upside down into

the inoculum and the vacuum was applied to each plant for three minutes. After inoculations, the plants were held at room temperature in the laboratory to await leaf sample collection. The plants were assessed for symptom development after sample collection.

Leaf samples were collected at 0, 8, and 15 hours post inoculations (hpi). All the above ground leaves were cut with sterilized scissors and immediately placed in liquid nitrogen after wrapping in aluminum foil. These time points were chosen to evaluate genes that change expression patterns early in response to *Cmn* infection and contribute to primary defense response. In a preliminary RT-PCR experiment designed to determine an optimal hpi of sampling, the pathogen responsive genes PR1 and PR5 were expressed within 12 to 15 hours. A previous study in Arabidopsis found that genes involved in early defense signaling responded to elicitor flg22 as early as 30 minutes after treatment (Asai et al 2002).

Control inoculations were done with water in place of the *Cmn* inoculum and samples were also collected from these controls at each time point. Three biological replicates with one plant representing a replicate were included for each line, treatment, and hpi combination, resulting in a total of 36 samples (Table 3.1). RNA was isolated using the Qiagen RNeasy mini kit (Qiagen, Valencia, CA, USA) based on the manufacturers suggested protocol, and was purified using the Qiagen RNA clean-up protocol according to the manufacturer instructions. RNA quality and quantity was determined using a Nanodrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE).

3.3.9 Library preparation and sequencing

RNA samples were submitted to the University of Minnesota Genomics Center for library preparation and sequencing. RNA sizing, quantification, and purity assessments were done with an Agilent Bioanalyzer (Agilent, Santa Clara, CA). Standard 36 dual indexed TruSeq RNA libraries were created. Libraries were sequenced on the Illumina HiSeq 2500 instrument using v4 chemistry (Illumina, Inc., San Diego, CA). Single end 50 bp reads were generated from the sequencing runs.

3.3.10 Sequence quality control and counting of reads

The quality of the reads was determined using FastQC version 0.11.5 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). High contents of Illumina Universal and TruSeq adapters were detected and these adapters were removed with CutAdapt version 1.8.1 (Martin, 2011). Quality cutoff of 20 and minimum processed read length of 20 were used in CutAdapt while processing the reads. After adapter trimming, the reads were aligned to the B73 v4 reference genome using a spliced aligner TopHat2 version 2.0.13 (Kim et al. 2013) using the following parameters: -i 5 -I 60000. The B73 v4 genome assembly was downloaded from Gramene Release 33 (http://www.gramene.org). Counting of reads was done with samtools version 1.2 and HTSeq version 0.5.3 (Anders et al., 2015). HTSeq was run with the following parameters: -s no -t gene -i ID -m union -a 20.

3.3.11 Gene co-expression network analysis

The raw expression matrix of 39,324 genes was filtered to remove genes with consistently low counts across the samples by keeping the genes with more than 10

counts in 90% of the samples. The filtered count matrix was read into the DESeq2 R package and a variance stabilizing transformation was used to normalize and transform the data using the varianceStabilizingTransformation function (Love et al., 2014). After applying the transformation, the expression matrix was processed with R package WGCNA to identify gene co-expression modules (Langfelder and Horvath, 2008). The WGCNA procedure involves the following steps: 1) Calculation of correlation matrix for all the genes from the gene expression matrix. The default standard Pearson correlation was used to obtain the correlation matrix of genes; 2) Calculation of adjacency matrix from the correlation matrix by raising the correlation matrix to soft power β. Soft threshold power β was chosen based on the results from pickSoftThreshold function, which performs a network topology analysis for a user defined set of candidate powers. A set of network indices was then plotted against different values of β. The plot of scale free topology index and βs was inspected and a $\beta = 20$ at which the fit index reached 0.90 was chosen; 3) Calculation of the topological overlap matrix (TOM), which was used to produce a hierarchical clustering dendrogram; 4) Identification of gene co-expression modules from the hierarchical cluster tree created from TOM by using a dynamic tree cut procedure. The function blockwiseConsensusModules inside the WGCNA package was used to identify the modules. Several values of the tree cut height parameter were tested, and a tree cut height of 0.25 provided a reasonable number of clean modules. Module eigen genes that are representative of expression profile of all the genes inside each module were visualized by a heat map to find any trends in expression profiles across the control and treated samples. The normalized and transformed expression matrix was

centered and standardized to calculate the Z-scores and the expression profile of all the genes inside each module across the control and treated samples was plotted using parallel coordinate plots. A gene ontology (GO) analysis was performed using the AgriGO server to test if any of the modules were enriched in functions related to defense (Du et al., 2010).

3.4 Results

3.4.1 Phenotypic variability among inbred lines

Inbred lines included in the diversity panel showed a skewed distribution towards resistance similar to what has been observed earlier for bi-parental mapping populations (Figure 3.1; Singh et al. 2016). While only a few plots showed severe wilting following inoculations, leaf blighting symptoms were frequently observed. In the analysis of variation on the diversity panel, main effect of inbred line, and interaction effect of inbred line-by-year were found to be significant. Plot-based heritability in the diversity panel was found to be high ($H^2 = 0.75$), indicating a combination of a large amount of genetic variation for resistance and high measurement precision.

3.4.2 Presence of population structure and rapid linkage disequilibrium decay

Two filters were applied to the SNP data before conducting PCA, LD decay, and GWAS analyses. Markers with more than 80% missing data and a minor allele frequency (MAF) less than 0.05 were filtered out, leaving 342,237 SNPs for analysis. LD decayed to 0.2 within 1000 bp, suggesting a high amount of diversity in the chosen panel. Principal component analysis using the SNP dataset revealed subpopulations within the diversity panel as seen from the plot of the first two PCs (Figure 3.2a). Principal

component 1 (PC1) separated the lines according to stiff stalk, non-stiff, and popcorn subpopulations. A few sweet corn and tropical lines included in the panel also clustered among themselves (Figure 3.2a). Ancestry estimates for K = 3 from ADMIXTURE were plotted to visualize the ancestral background of each line (Figure 3.2b). Stiff stalk, non-stiff stalk, and unclassified lines consisted of mixed ancestries, and unclassified lines showed similar ancestry to non-stiff stalk lines. Popcorn lines mostly had one ancestral type, which was distinct from the stiff stalk and non-stiff stalk types (Figure 3.2b).

A neighbor joining tree of the diversity panel lines was created to visualize the subpopulations and distribution of Goss's wilt resistance by subpopulation. Color coding of lines by the subpopulations clearly indicated that the lines clustered according to the subpopulations as expected (Figure 3.3a). No strong pattern in the distribution of lines according to Goss's wilt resistance/susceptibility was observed when the lines were color coded by their disease score (Figure 3.3b). In other words, clusters of related lines contained both resistant and susceptible lines, and Goss's resistance did not seem to correspond to any particular subpopulation. This pattern is indicated by random spread of green (resistant) and red (susceptible) colors on the neighbor joining tree (Figure 3.3b). Although elite popcorn breeding lines are generally highly susceptible to Goss's wilt (Rocheford *et al.* 1985; O. Rodriguez, personal communication), the popcorn lines screened as part of this diversity panel ranged from resistant to highly susceptible, indicating a high degree of variation even within popcorn (Figure 3.3b). We did observe that lines closely related to Mo17 were generally resistant, and most of the lines related to

B14 were moderate to highly susceptible as has been found in previously (Calub et al., 1974a).

Although the plot of PC1 and PC2 showed clustering of lines according to subpopulations (Figure 3.2a), the amount of phenotypic variation explained by the first three components as calculated by fitting them in a linear model with Goss's wilt BLUEs as phenotype, was only 3% in the diversity panel dataset. In the combined dataset of the diversity panel and the bi-parental families, the population structure explained 7% of the phenotypic variation for Goss's wilt.

3.4.3 Association analyses identified SNPs with small effects

Association analysis within the diversity panel (555 lines) identified three SNPs on chromosome 5 that passed a FDR cutoff of 0.10 (Table 3.2, Figure 3.4a). The amount of phenotypic variation for Goss's wilt resistance explained by each SNP in the diversity panel ranged from 3% to 5% (Table 3.2). The three significant SNPs together explained 8% of the phenotypic variation for Goss's wilt resistance after accounting for the variation explained by the population structure. The genomic kinship matrix explained 64% variation for Goss's wilt resistance, which is 50 fold higher than the variation explained by significant SNPs only. In the association analysis of the combined dataset (diversity panel and RIL populations), 10 SNPs were found to be significant at a FDR of 0.10. These SNPs were on chromosomes 1, 2, and 5 (Table 3.2, Figure 3.4b). A cluster of significant SNPs (five) was detected on chromosome 2. Chromosome 1 had three significant SNPs, and chromosome 5 had two significant SNPs associated with Goss's wilt resistance. Each of the SNPs identified in the combined dataset explained a small

amount of phenotypic variation for Goss's wilt resistance, ranging from 1% to 3%. Physical positions of the significant SNPs in QTL mapping of bi-parental families conducted in Singh et al. (2015) and GWAS in the combined dataset were plotted for each of the chromosomes on which significant associations were detected. Significant peaks in GWAS on chromosomes 1, 2, and 5 co-localized with QTL peaks of the bi-parental QTL mapping (Figure 3.5).

3.4.4 Assessing haplotype allele frequency in sub populations

A haplotype block analysis at each of the genomic regions harboring significant SNPs revealed two types of haplotype block structure. The haplotype blocks at the chromosome 1 and 5 regions were very small in size as a result of low LD in these regions, while chromosome 2 had larger haplotype blocks indicating limited recombination in this region (Figure 3.6a, 3.6b). A 10 kb region around a significant SNP on chromosome 5 consisted of five haplotype blocks with a maximum haplotype block length of 4 kb for block 2 (Figure 3.6b). In contrast, on chromosome 2, a region with five significant SNPs consisted of four haplotype blocks of length 5 kb, 79 kb, 53 kb, and 13 kb. Out of five significant SNPs on chromosome 2; four SNPs were located on block 1 (5 kb). The fifth SNP was 1,977,875 bp away from this haplotype.

The chromosome 2 region with significant SNPs and larger haplotype blocks was investigated further to look at the haplotype allele frequency in the diversity panel as well as in subpopulations within the panel. Block 1, 2, 3, and 4 of the chromosome 2 haplotype consisted of 3, 6, 3, and 4 alleles, respectively (Table 3.3). Four out of five significant SNPs identified on chromosome 2 were located in block 1. Allele A1 was a

resistant allele, meaning that the mean rating of the lines carrying this allele was lower than the allele A2 and A3. Resistant allele (A1) was frequent in all sub populations but infrequent in popcorn (2.0% of lines). Moderately susceptible allele A2 had a very high frequency in popcorn (93.8% of lines). Interestingly, the most susceptible allele, A3 had a relatively high frequency in stiff stalk (29% of lines) and non-stiff stalk (20% of lines) sub populations but had low frequency in popcorn sub population (4.2% of lines). This result indicates that there is an allele in stiff stalk and non-stiff stalk sub populations contained in some lines that confers higher susceptibility than the popcorn allele (Table 3.3). Each haplotype allele, however, had a small effect on mean disease as the mean disease of the lines carrying each of the haplotype allele ranged from 2.28 - 2.97 (Table 3.3).

3.4.5 Candidate genes

Potential candidate genes underlying the significant SNPs were identified by aligning the physical positions of the SNPs to the B73 RefGen_v3 (version three) reference genome (MaizeGDB). To find the candidate genes, a window around the significant SNPs was defined by sliding both upstream and downstream until the LD decayed to 0.2. Using the genome browser, several protein coding genes were identified within these windows with notable functions such as nucleic acid binding, zinc ion binding, electron transport, and a Glutamate receptor protein (Table 3.2).

3.4.6 RNA sequencing, read numbers and quality

Standard dual indexed TruSeq RNA libraries were created from 36 samples. All the libraries were combined into a single pool and sequenced across four lanes of a flow

cell on the Illumina HiSeq 2500 instrument to generate more than 220 million reads on each lane. The reads were of high quality with a mean PHRED quality scores of greater than 20. Number of reads obtained for each of the samples was as targeted at approximately 20 million. Illumina universal and TruSeq adapter sequences were detected in a few samples and were removed using CutAdapt (Martin, 2011). Mapping of reads to the B73 reference genome version 4 revealed that four samples had overrepresented sequences as a larger proportion of the reads for these samples mapped to multiple locations. These samples were filtered out before further processing.

3.4.7 Two gene modules responded to inoculations with *Cmn*

To further refine and identify candidate genes, a time course experiment of a resistant (N551) and susceptible (B14A) inbred line under control and infected conditions was done. Weighted gene co-expression network analysis identified 15 modules of co-expressed genes in the time course experiment. The number of genes inside each module ranged from 75 in module 15 to 3513 in module 1, and 3840 genes did not cluster with any other genes (Table 3.4). A heatmap of the eigen genes of the modules provide a high-level overview, allowing identification of trends in gene expression changes of possible biological importance. Two modules were identified that may be associated with *Cmn* infection: modules 13 and 15. Module 13 showed changes in gene expression in B14A at 15 hpi but no changes in expression were apparent in N551. Module 15 showed an increase in gene expression levels in resistant line N551 at 8 and 15 hpi with *Cmn* (Figure 3.7). Further evaluation of genes inside module 13 from parallel coordinate plots revealed that most genes were up-regulated in B14A at 15 hpi with *Cmn* (Figure 3.8A). A set of

anti-correlated genes in module 13 showed decreased expression at 15 hpi with *Cmn*. Genes inside module 15 showed a three-way interaction that involved inbred line N551, treatment, and time (Figure 3.8B). Most of the genes inside module 15 increased expression in response to *Cmn* in resistant line N551. Six genes inside module 15 also showed a decreased expression at 8 and 15 hpi in response to *Cmn.* These changes in expression pattern were not observed in susceptible line B14A.

3.4.8 Module 13 was enriched in genes with defense related functions

Gene ontology (GO) enrichment analysis was done to see if the genes comprising modules 13 and 15 were enriched in specific functions related to defense. Thirty-six percent of the characterized genes that comprised module 13 have defense-related GO terms. Significant biological process related GO terms that have defense related functions included death, cell death, programmed cell death, and apoptosis. Significant molecular function related GO terms for the genes in module 13 that may have a role in defense included kinase activity, protein kinase activity, protein serine/threonine activity, and protein tyrosine activity. All the significant GO terms for module 13 are shown in Table 3.5. Checking the functional annotations of genes inside module 13 revealed 8 genes with serine/threonine kinase function, a gene with calmodulin binding domain, a NBS-LRR type disease resistance protein, a pathogen related protein 5 (PR-5), and a disease resistance gene analog.

Genes of the module 15 did not show any apparent GO enrichment related to defense. Significant molecular function related GO terms associated with genes in module 15 included catalytic activity, lyase activity, cofactor binding, carbon-carbon

lyase activity, and carboxy-lyase activity. Significant biological process related GO terms for genes in module 15 were cellular amino acid and derivative metabolic process, and cellular catabolic process. Although the genes in module 15 were not enriched in defense related functions, when the annotations of these genes were checked manually, one of the genes was found to be annotated as pathogenesis related gene, PR-4.

3.4.9 Variation explained by genes inside modules for Goss's Wilt

We wanted to test whether the SNPs inside the genes included in modules 13 and 15 explained a disproportionate amount of the variance in Goss's wilt resistance observed in the diversity panel to determine if in fact these genes are important contributors to genetic variance for quantitative disease resistance. The physical positions of the GBS SNPs used in the GWAS analyses were aligned with the physical positions of the genes that clustered into modules 13 and 15. The SNPs inside these genes were used to calculate kinship matrices which were fit into a linear G-BLUP model to calculate the amount of variation that could be explained for Goss's wilt resistance using said kinship matrix. To determine if the variance explained is enriched by the module 13 and 15 kinship matrices, additional kinship matrices were calculated with random SNPs equal in number to the SNPs for modules 13 and 15, which were 392 and 55 SNPs, respectively. The random SNPs were randomly sampled 100 times. The SNPs in module 13 and 15 explained 18.5% and 10.8% of the phenotypic variation for Goss's wilt resistance, respectively. The percent variation explained by the same number of random SNPs as in module 13 varied from 16.4% to 28.9%. Only 11 out of 100 random samples explained less than 18.5 % variation. This result shows that the percent variation explained by SNPs

in module 13 fell within the range and was not significantly higher than the variation explained by random sets of SNPs. The percent variation explained by same number of random SNPs as in module 15 ranged from 4.5% to 13.7%, indicting no enrichment in variance explained by SNPs in module 15 also.

3.5 Discussion

3.5.1 General findings

Goss's wilt is an important disease of maize that had not been studied extensively with the use of current molecular technologies as compared to other maize diseases such as NCLB, SCLB, and gray leaf spot (Poland et al., 2011; Kump et al., 2011; Benson et al., 2015). Only two studies have been reported to use molecular markers to study the genetic basis of Goss's wilt resistance (Schaefer and Bernardo, 2013; Singh et al., 2016). In this study a combination of two approaches -- genome-wide association mapping and gene co-expression network analysis -- were used to increase our knowledge of the genetic basis of resistance to Goss's wilt of maize. The two approaches used in this study pointed towards a complex nature of resistance to Goss's wilt of maize, which is in agreement with the complex nature of plant signaling and defense processes as studied using model organisms (Jones and Dangl, 2006; Bonardi and Dangl, 2012) and hypothesized to be the case for Goss's wilt based on initial genetic studies (Martin et al., 1975; Rocheford et al., 1989; Treat et al., 1990). A complex, two level immune system is in place in plants to detect and defend themselves against pathogens which comprises pathogen associated molecular pattern (PAMP) triggered immunity (PTI) and effector triggered immunity (ETI). While PTI is the basal level of defense provided by membrane

receptors such as receptor like kinases, ETI is triggered by pathogen avirulence factors and involves nucleotide binding leucine rich repeat proteins (NB-LRR) (Jones and Dangl 2006).

The GWAS mapping performed in this study supported the findings of the previous linkage mapping study (Singh et al., 2016). By combining data from bi-parental families and the diversity panel, we were able to identify loci associated with Goss's wilt at a higher resolution as compared to the low resolution obtained by just using the bi-parental families. Besides identifying SNPs associated to Goss's wilt resistance, we characterized a large and diverse panel of maize lines for Goss's wilt resistance and assessed the effect of population structure on resistance. Gene co-expression network analysis was conducted to explore major modules of genes that may be involved in response of resistant and susceptible maize inbred lines to *Cmn*. The gene co-expression network analysis conducted as a part of this study is the first large scale transcriptome analysis of the response of resistant and susceptible maize inbred lines to *Cmn*.

3.5.2 Phenotypic distribution of Goss's wilt resistance in maize germplasm

The distribution of Goss's wilt resistance within the diversity panel was skewed towards resistance. Our expectation was that the inbred lines within certain sub-populations would be resistant as compared to inbred lines from other sub-groups. This was, however, not observed in this study as resistant and susceptible lines were distributed among all the groups (Figure 3.3b). Schuster *et al.* (1972) and Calub *et al.* (1974) tested over 100 maize inbred lines from different genetic backgrounds for resistance to Goss's wilt and found that the reaction of the inbred lines varied from highly

resistant to highly susceptible. Similar to what has been reported in historical studies (Schuster et al., 1972a; Calub et al., 1974a), the lines genetically related to inbred line B14 were highly susceptible, and the lines related to Mo17 were found to be resistant in the present study.

3.5.3 Significant SNP associations

Significant SNPs were detected on chromosome 5 in the diversity panel dataset and on chromosomes 1, 2, and 5 in the combined dataset of the diversity panel and bi-parental families. Similar to our previous joint-linkage mapping study, each of the significant SNPs associated with Goss's wilt resistance in the GWAS explained small amount of the phenotypic variation for Goss's wilt resistance. This result is in accordance with the results obtained for other important leaf diseases of maize including southern corn leaf blight, northern corn leaf blight, and gray leaf spot for which small effect SNPs were associated with resistance (Kump et al., 2011; Poland et al., 2011; Benson et al., 2015). Thirty-two QTL with small additive effects that together explained 80% and 93% of the phenotypic and genotypic variation respectively for southern corn leaf blight were reported in a GWAS of maize NAM population (Kump et al., 2011). Similarly, 29 QTL were identified for northern corn leaf blight in NAM population that explained 77% and 96% of the phenotypic and genetic variance respectively for northern corn leaf blight, with each QTL being of small effect (Poland et al., 2011). The amount of total phenotypic variation explained by significant SNPs for Goss's wilt resistance however was much lower than that explained by QTL for southern corn leaf blight and northern corn leaf blight in previous studies because of much larger number of QTL detected using

the entire NAM population than a few significant SNPs detected using the diversity panel for Goss's wilt resistance.

A small amount of phenotypic variation (10% in diversity panel and 17% in combined dataset) explained by significant SNPs can be attributed to the combined effect of two situations discussed below. As explained by Yang *et al.* (2010) for human height, either the causal SNPs that explain very small amount of variation do not pass the significance in GWAS, or there is a lack of complete LD between the causal variants and the genotyped SNPs. A genomic kinship matrix explained 64% and 46% of the phenotypic variation for Goss's wilt resistance in the diversity panel and combined dataset respectively. The high proportion of phenotypic variation explained by the kinship matrix indicates that the genetic architecture of Goss's wilt resistance is complex and that genomic selection approach may be more effective in breeding for Goss's wilt resistance compared to selecting for individual QTL. A similar conclusion was reported in a *Fusarium* ear rot study in maize in which the kinship matrix explained nearly 50% of the variation for *Fusarium* ear rot and only 1.3 - 3% of the variation was explained by individual significant SNPs (Zila, 2014).

3.5.4 Candidate genes from GWAS

Due to the high mapping resolution often achieved, especially in species such as maize, GWAS mapping results can pinpoint potential candidate genes. Several candidate genes underlying significant SNPs were identified in this study with a range of functional annotations (Table 3.2). On chromosome 5, a gene coding for Glutamate receptor protein was identified. The glutamate receptor-like gene family (*GLRs*) is homologous to the

mammalian ionotropic glutamate receptor gene family. In plants, *GLRs* have been reported to play an important role in plant defense response by coding for amino acids that act as sensors to detect pathogen attack (Forde and Roberts, 2014). Overexpression of radish GLR in Arabidopsis has been reported to lead to increased resistance against the fungus *Botrytis cinerea* and enhanced expression of defense genes (Kang et al., 2006). Similarly, knock out mutants of *AtGLR* in Arabidopsis showed increased susceptibility to bacteria *Pseudomonas syringae*. Activation of defense genes expression was also dependent upon *AtGLR* (Li et al., 2013).

The mechanism of quantitative disease resistance (QDR) is complex and probably involves a diverse range of genes. The mechanism of QDR has been poorly understood as compared to qualitative resistance (Poland et al., 2009). Poland et al. (2009) proposed six hypotheses that may contribute to QDR: (i) genes contributing to plant development and morphology may have pleiotropic effects on QDR, (ii) receptor like kinase genes that play a role is PAMP triggered immunity may contribute to QDR, (iii) defense enzymes secreted by the plants to detoxify phytotoxins may have a role in QDR, (iv) genes that regulate salicylic acid, jasmonic acid, and ethylene pathways involved in transduction of defense signals may affect QDR, (v) attenuated R-genes that provide partial resistance may be involved in QDR, and (vi) QDR is conditioned by unknown genes with entirely different function. The quantitative nature of Goss's wilt resistance indicated that the genes involved in multiple hypotheses proposed above may collectively contribute to resistance.

3.5.5 Weighted gene co-expression network analysis

The number of host genes regulated by disease resistance is a basic systems biology question that has been explored multiple times with different plant pathosystems. Using interlog and domain based computational approaches 2,043 Arabidopsis proteins were predicted to interact with bacterial pathogen *Pseudomonas syringae* (Sahu et al., 2014). In a gene-coexpression network analysis of response of Citrus to bacterium *Candidatus Liberibacter* spp., 3,507 genes were suggested to play a role in defense (Zheng and Zhao, 2013). Similarly, gene co-expression network analysis of genes regulated by immune and hypersensitive resistance responses to *Blumeria graminis* f. sp *tritici* in diploid wheat *Triticum urartu* indicated that 3,900 and 4,100 genes may be involved in the above two types of resistance responses respectively (Zhang et al., 2016). In the present study, two modules (module 13 and 15) that corresponded to response of maize inbred lines to *Cmn* were identified consisting of 318 and 75 co-related genes, respectively.

Most of the genes in module 13 showed increased expression in susceptible inbred line B14A at 15 hpi and apparently no up-regulation in resistant line N551. This is an interesting result as previous findings from molecular and co-expression network studies have shown that plant immunity is controlled by negative regulation of a certain set of genes (Sato et al., 2010; Segonzac et al., 2014). Particularly, specific subunits of protein Ser/Thr phosphatases have been involved in negative regulation of defense signaling at different steps (Segonzac et al. 2014). Negative regulation of certain genes within a plant immune system are required for the effective functioning by preventing

over activation of the immune system which may cause auto immune responses such as cell death thus reducing the plant fitness (Sato *et al.* 2010). The GO analysis of module 13 revealed significant GO terms for 56 phosphorous metabolic process related genes which may explain why the genes inside module 13 were not up regulated in resistant line N551. Both in module 13 and 15, at least 50% of the genes are either annotated as hypothetical proteins or putative uncharacterized proteins. Future functional annotation of these genes may be informative on what other type of genes are involved in the *Cmn*-maize interaction.

We hypothesized an overlap between the candidate genes identified in GWAS and the genes in modules that responded to *Cmn* in gene-co-expression network analysis. However, none of the eight candidate genes overlapped with the genes inside modules 13 and 15. This result can be explained by low power of GWAS such that it cannot detect all of the phenotypic variation for a given trait due to low frequency of the causal alleles that cannot be tagged by genotyped SNPs and rapid LD decay in this panel (Wray et al., 2013).

The genes implicated in the WGCNA did not explain a disproportionate amount of variance in Goss's wilt resistance in the diversity panel using a G-BLUP model. This could be because the variation in gene expression of these genes is not related to polymorphisms within the genes, but rather to polymorphisms in cis- or trans-acting transcription factors. Also, it is possible that the gene expression differences between the two chosen inbred lines, B14A and N551, are not representative of the gene expression patterns in the diversity panel as a whole. Thirdly, the genes contributing to quantitative

disease resistance in the field 15 days or more after inoculation may not be the same, or have little overlap with, genes responding to infection within hours during the seedling stages. Finally, while a mixture of five isolates was used for field inoculations of the GWAS, only one isolate 12038 was used to inoculate B14A and N551 in the lab. Isolate 12038 was not a part of the mixture of five isolates used for field inoculations.

3.6 Conclusions

In conclusion, the association mapping identified SNPs associated with resistance to Goss's wilt of maize. Each of the loci identified explained a relatively small amount of phenotypic variation for Goss's wilt resistance. Haplotype alleles in the region around the significant SNPs had a very small effect on Goss's wilt resistance. These results are consistent with the findings of our previous study conducted using bi-parental families. Given the rapid LD decay in the diversity panel, increasing the number of markers and population size may help to detect more loci associated with Goss's wilt resistance. Gene co-expression network analysis identified the genes that may be regulated by the response of maize inbred lines to *Cmn*. The modules of correlated genes that respond to *Cmn* provided an important information about the number and type of genes involved in resistance to Goss's wilt. This is the first gene expression study and in combination with GWAS, it increased our understanding of Goss's wilt, a disease that is not highly researched so far. For breeding maize for resistance to Goss's wilt however, due to lack of any major effect QTL in the germplasm used in this study and laborious efforts required in phenotyping, genomic selection approach should be explored.

Table 3.1 Design of RNA sequencing experiment. Each row of the table displays combination of treatment including inbred line (B14A or N551), inoculation medium (water or *Cmn* bacteria), hours post inoculation (hpi) at which the samples were taken for RNA isolations (0, 8, 15 hpi), and if the sample was kept or filtered out before conducting weighted gene co-expression network analysis (WGCNA).

Line	Treatment	Hours post inoculation	Included/not included in WGCNA
B14A	Water	8	No
B14A	Water	8	Yes
B14A	Water	8	Yes
B14A	Water	15	Yes
B14A	Water	15	Yes
B14A	Water	15	No
B14A	CMN	0	Yes
B14A	CMN	0	Yes
B14A	CMN	0	Yes
B14A	CMN	8	Yes
B14A	CMN	8	Yes
B14A	CMN	8	Yes
B14A	CMN	15	Yes
B14A	CMN	15	Yes
B14A	CMN	15	Yes
N551	Water	8	Yes
N551	Water	8	Yes
N551	Water	8	Yes
N551	Water	15	Yes
N551	Water	15	No
N551	Water	15	Yes
N551	CMN	0	No
N551	CMN	0	Yes
N551	CMN	0	Yes
N551	CMN	8	Yes
N551	CMN	8	Yes
N551	CMN	8	Yes
N551	CMN	15	Yes
N551	CMN	15	Yes
N551	CMN	15	Yes

Table 3.2 Physical positions of the SNPs significantly associated with Goss's wilt resistance in the diversity panel and combined dataset.

Chrom	Physical Positon of SNP (bp)	p-value	q-value	R^2	SNP effect	Genes Underlying Significant SNPs	Gene Function
			Diversity panel				
5	210,554,445	4.25×10^{-7}	0.10	0.03	-0.26	GRMZM2G368206	PHD finger domain Zinc ion binding
5	210,554,466	4.25×10^{-7}	0.10	-	-0.26		
5	46,455,199	7.01×10^{-8}	0.02	0.05	0.28		Glutamate receptor
			Combined dataset (diversity panel and bi-parental families)				
1	182,307,976	9.49×10^{-7}	0.09	0.02	-0.23		
1	182,307,992	2.34×10^{-6}	0.08	0.02	-0.22		
1	187,675,076	1.59×10^{-6}	0.09	0.01	-0.20	GRMZM2G132704	Nucleotide/RNA binding
						GRMZM2G132607	Ribokinase activity
						GRMZM2G132623	Constituent of ribosome
2	198,101,869	1.72×10^{-6}	0.09	0.01	0.20	GRMZM2G048582	Response to Nitrogen
						GRMZM2G048551	Zn ion binding
						GRMZM2G512469	Unknown
2	198,101,827	3.18×10^{-6}	0.09	0.01	0.20		
2	198,101,829	3.18×10^{-6}	0.09	-	0.20		
2	198,101,830	3.18×10^{-6}	0.09	-	0.20		
2	200,227,875	1.86×10^{-6}	0.09	0.01	0.21		
5	210,554,445	1.07×10^{-6}	0.09	0.03	-0.19	GRMZM2G368206	Protein binding, zinc ion binding
5	210,554,466	1.07×10^{-6}	0.09	-	-0.19		

Table 3.3 Haplotype allele frequency in the diversity panel and sub populations (stiff stalk, non-stiff stalk, popcorn, and unclassified) observed in the chromosome 2 region. Mean Goss's wilt ratings of the lines carrying each haplotype allele are presented.

Block	Haplotype Allele	Allele No	Haplotype Allele Frequency					Mean Rating
			Panel	Stiff	Non-stiff	Popcorn	Unclassified	
BLOCK1 (5 kb)	ATGG	A1	0.651	0.657	0.79	0.021	0.722	2.38
	AGGT	A2	0.165	0.051	0.01	0.938	0.1	2.73
	CGAG	A3	0.184	0.293	0.2	0.042	0.178	2.89
BLOCK2 (79 kb)	CCTCAGAAACCACGGCGGA	B1	0.379	0.474	0.438	0	0.465	2.35
	TTCTGGGATATAAAGCCTA	B2	0.085	0.01	0.01	0.467	0.069	2.85
	TTCTGAGATCCGCGCCGTA	B3	0.158	0.278	0.177	0.044	0.171	2.92
	CCTCAGATACCACGGCGTC	B4	0.113	0.062	0.219	0	0.139	2.47
	CCTCAGATACCACGGTGTC	B5	0.097	0.124	0.156	0	0.098	2.28
	TTTCGGGATCCGCGGCGTA	B6	0.081	0.052	0	0.489	0.057	2.57
BLOCK3 (53 kb)	GCC	C1	0.463	0.511	0.303	0.478	0.5	2.44
	ACC	C2	0.175	0.283	0.182	0.043	0.172	2.97
	GTT	C3	0.363	0.207	0.515	0.478	0.328	2.44
BLOCK4 (13 kb)	CGGACG	D1	0.567	0.811	0.531	0.043	0.63	2.53
	TATACG	D2	0.113	0.078	0.136	0.109	0.103	2.64
	TATACC	D3	0.225	0.044	0.309	0.37	0.217	2.56
	TATGGG	D4	0.095	0.067	0.025	0.478	0.049	2.50

Table 3.4 Number of genes in each of the 15 modules identified through weighted gene co-expression analysis (WGCNA).

Module	Number of Genes
1	3513
2	3289
3	3155
4	2355
5	1132
6	971
7	640
8	616
9	586
10	583
11	390
12	331
13	318
14	311
15	75

Table 3.5 Significant gene ontology terms associated with the genes in module 13 identified from GO enrichment analysis.

GO term	Description	Number of genes in input list	Number of genes in BG/Ref	p-value	FDR
GO:0006468	protein amino acid phosphorylation	54	1528	4.9×10^{-20}	2×10^{-17}
GO:0016310	phosphorylation	55	1821	1.9×10^{-17}	2.6×10^{-15}
GO:0043687	post-translational protein modification	55	1832	2.4×10^{-17}	2.6×10^{-15}
GO:0006464	protein modification process	58	2013	1.6×10^{-17}	2.6×10^{-15}
GO:0043412	macromolecule modification	58	2069	5.4×10^{-17}	4.6×10^{-15}
GO:0006796	phosphate metabolic process	56	1969	1.2×10^{-16}	7.2×10^{-15}
GO:0006793	phosphorus metabolic process	56	1969	1.2×10^{-16}	7.2×10^{-15}
GO:0044267	cellular protein metabolic process	62	3535	2.4×10^{-9}	1.3×10^{-7}
GO:0019538	protein metabolic process	70	4516	2.6×10^{-8}	1.2×10^{-6}
GO:0006915	apoptosis	9	129	1.9×10^{-6}	7.4×10^{-5}
GO:0012501	programmed cell death	9	129	1.9×10^{-6}	7.4×10^{-5}
GO:0008219	cell death	9	149	5.9×10^{-6}	0.00019
GO:0016265	death	9	149	5.9×10^{-6}	0.00019
GO:0044260	cellular macromolecule metabolic process	83	7062	9.4×10^{-5}	0.0028
GO:0051704	multi-organism process	7	132	0.00014	0.004
GO:0043170	macromolecule metabolic process	91	8085	0.00017	0.0046
GO:0044238	primary metabolic process	108	10134	0.00024	0.0059
GO:0044237	cellular metabolic process	95	9105	0.0019	0.044
GO:0004713	protein tyrosine kinase activity	53	1313	3.7×10^{-22}	1.1×10^{-19}
GO:0004674	protein serine/threonine kinase activity	53	1430	1.5×10^{-20}	1.7×10^{-18}
GO:0004672	protein kinase activity	55	1552	1.8×10^{-20}	1.7×10^{-18}
GO:0016773	phosphotransferase activity, alcohol group as acceptor	55	1824	2×10^{-17}	1.5×10^{-15}
GO:0016301	kinase activity	55	1909	1.4×10^{-16}	8.2×10^{-15}
GO:0005524	ATP binding	72	3193	2.5×10^{-16}	1.1×10^{-14}
GO:0032559	adenyl ribonucleotide binding	72	3197	2.7×10^{-16}	1.1×10^{-14}
GO:0030554	adenyl nucleotide binding	73	3384	1.4×10^{-15}	4.3×10^{-14}
GO:0001883	purine nucleoside binding	73	3384	1.4×10^{-15}	4.3×10^{-14}
GO:0001882	nucleoside binding	73	3385	1.5×10^{-15}	4.3×10^{-14}
GO:0032555	purine ribonucleotide binding	73	3572	2.3×10^{-14}	5.7×10^{-13}
GO:0032553	ribonucleotide binding	73	3572	2.3×10^{-14}	5.7×10^{-13}
GO:0000166	nucleotide binding	77	3970	4.9×10^{-14}	1.1×10^{-12}
GO:0017076	purine nucleotide binding	74	3767	1.1×10^{-13}	2.2×10^{-12}

GO:0016772	transferase activity, transferring phosphorus-containing groups	55	2285	2.2×10^{-13}	4.4×10^{-12}
GO:0016740	transferase activity	66	3613	1×10^{-10}	1.9×10^{-9}
GO:0003824	catalytic activity	125	11249	2.4×10^{-6}	4.1×10^{-5}
GO:0005529	sugar binding	9	177	2.2×10^{-5}	0.00035
GO:0005509	calcium ion binding	12	352	4.4×10^{-5}	0.00068
GO:0030246	carbohydrate binding	9	246	0.00024	0.0036

Figure 3.1 Histogram showing the distribution of weighted mean disease scores for inbred lines included in the diversity panel. Vertical dotted lines display the check lines B73 and B14A.

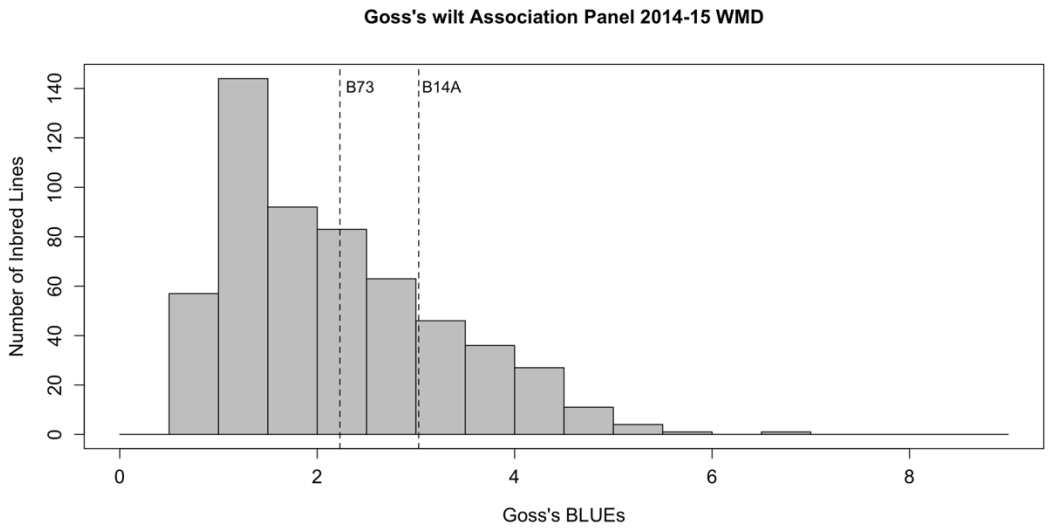
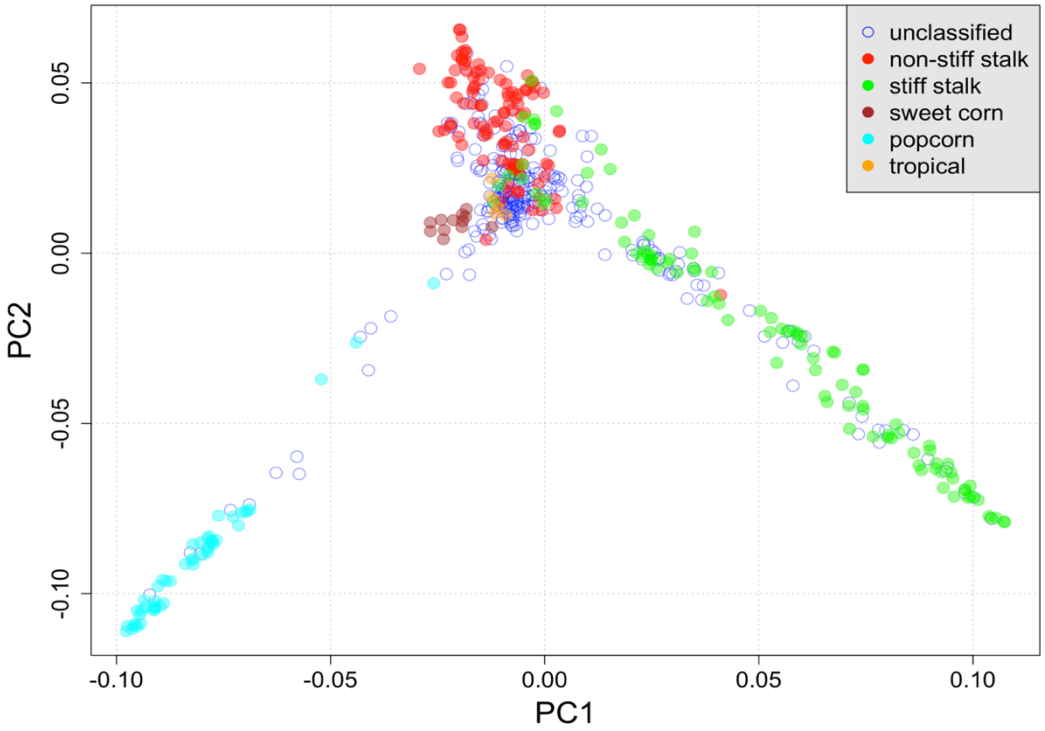


Figure 3.2 Population structure within the diversity panel. Population structure was assessed using (a) plot of the first two principal components (PCs) calculated from a principal component analysis on the SNP matrix of the diversity panel. A priori subpopulation classifications were according to Romay et al. (2013), (b) Population structure within the diversity panel assessed using ADMIXTURE. Inbred lines are arranged according to a priori subpopulation classifications take from Romay et al. (2013). Three clusters are displayed. Each bar of the plot represents a single inbred line with its proportional assignment to each cluster represented by the relative length of each color.



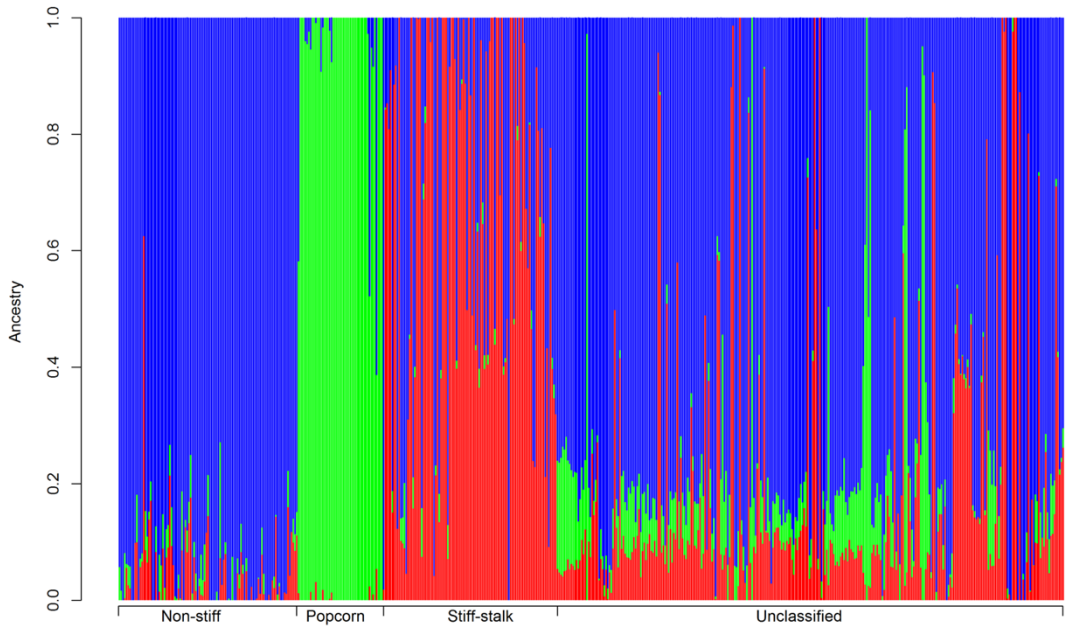
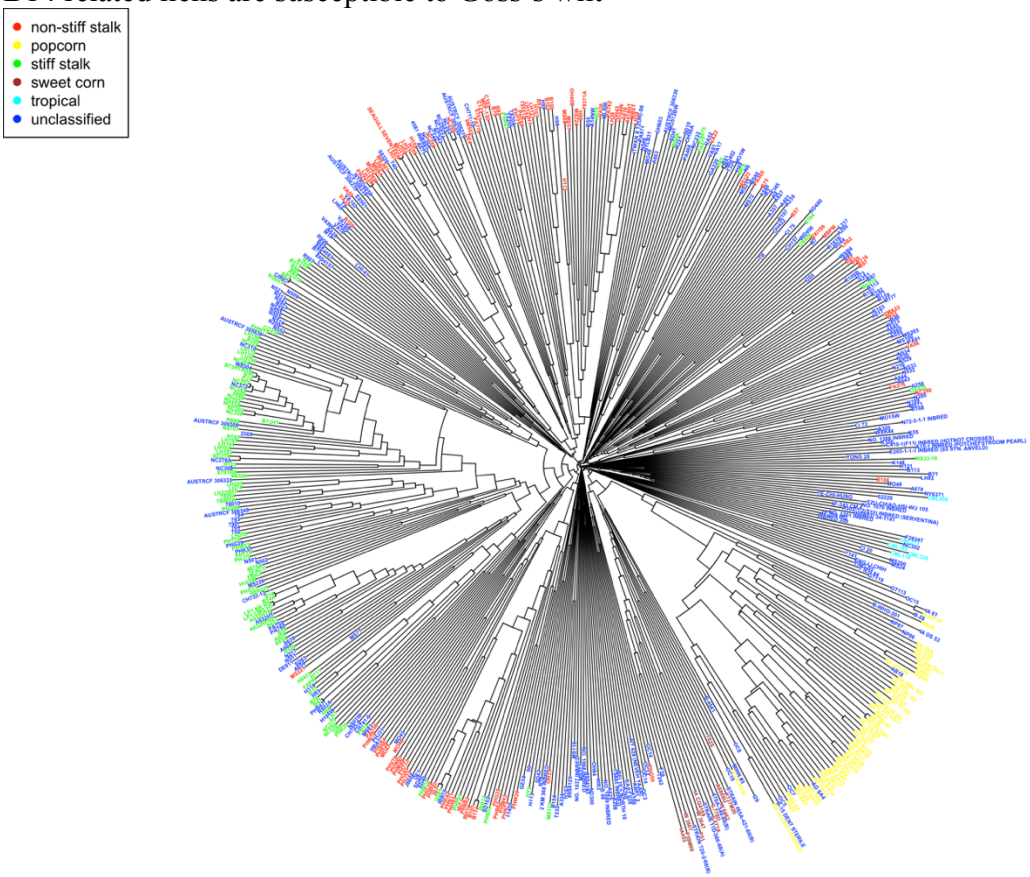

Ancestry
0.0
0.2
0.4
0.6
0.8
1.0
Non-stiff
Popcorn
Stiff-stalk
Unclassified

Figure 3.3 Neighbor joining tree of 555 lines of the diversity panel created from the distance matrix calculated with the SNP data. (a) The color coding of the edge labels indicates the six subpopulation groups within the diversity panel, (b) Red color indicates that a line is susceptible and green color represents resistant lines. Specific groups are zoomed to indicate trends in Goss's wilt resistance distribution. For example, most of the B14 related liens are susceptible to Goss's wilt



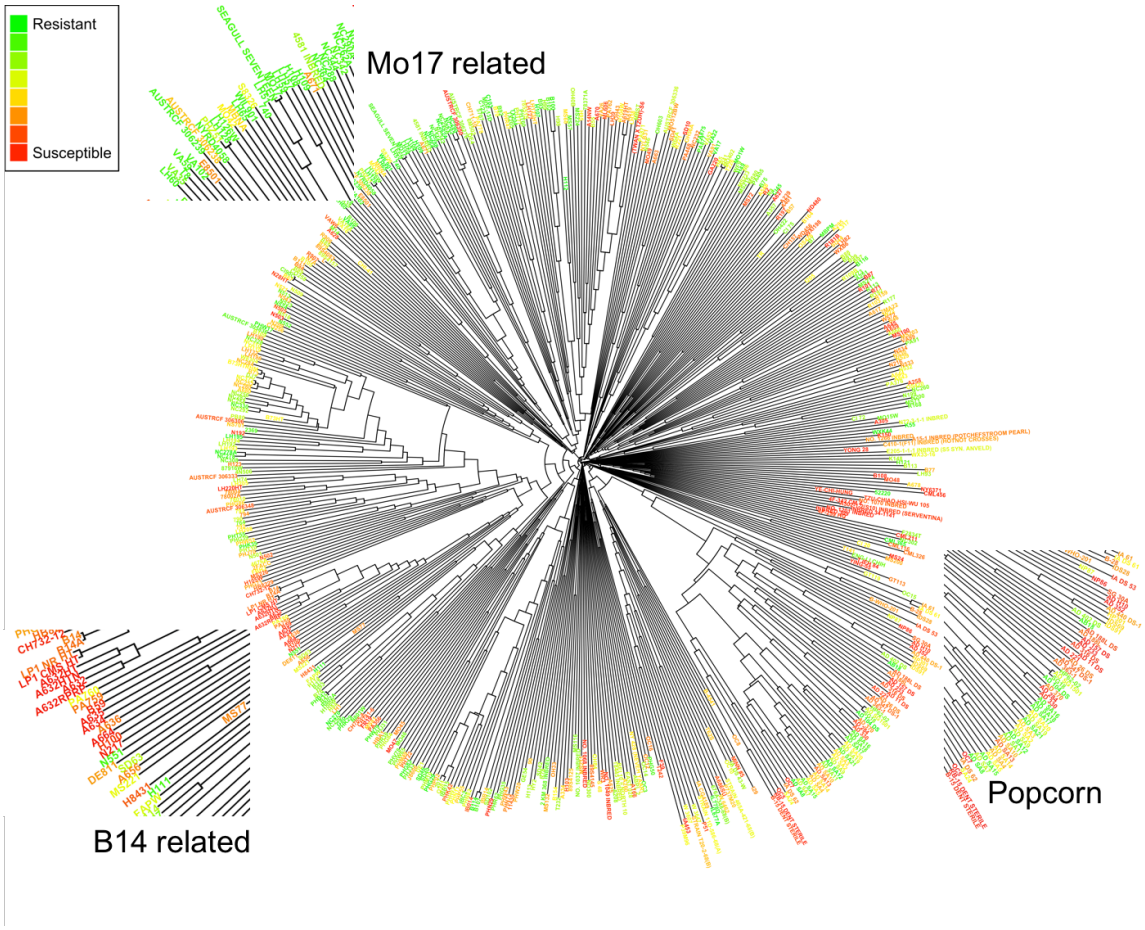
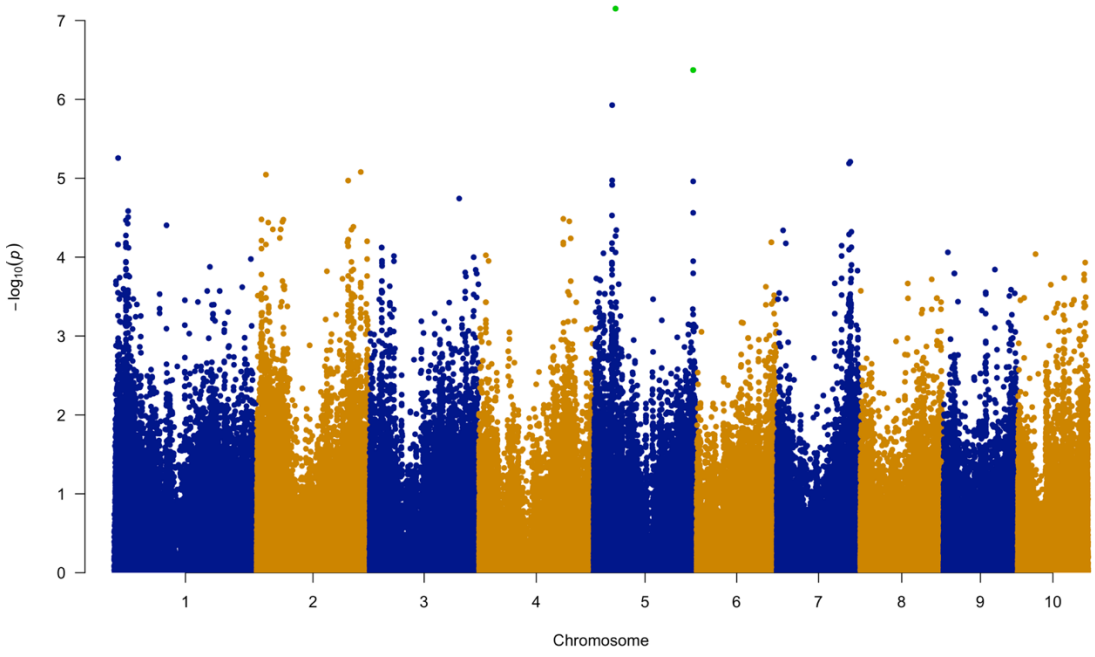

Resistant
Susceptible
Mo17 related
B14 related
Popcorn

Figure 3.4 Genome-wide association mapping results. (a) Results on the diversity panel (N =555) displayed in a Manhattan plot where the y-axis is the negative $\log_{10}$ of p-values for the SNPs from model [2]. Associations that passed the false-discovery rate of 0.10 are colored green, (b) results on the combined dataset (N =1005) displayed in a Manhattan plot where the y-axis is the negative $\log_{10}$ of p-values for the SNPs from model [2]. Associations that passed the false-discovery rate of 0.10 are colored green.



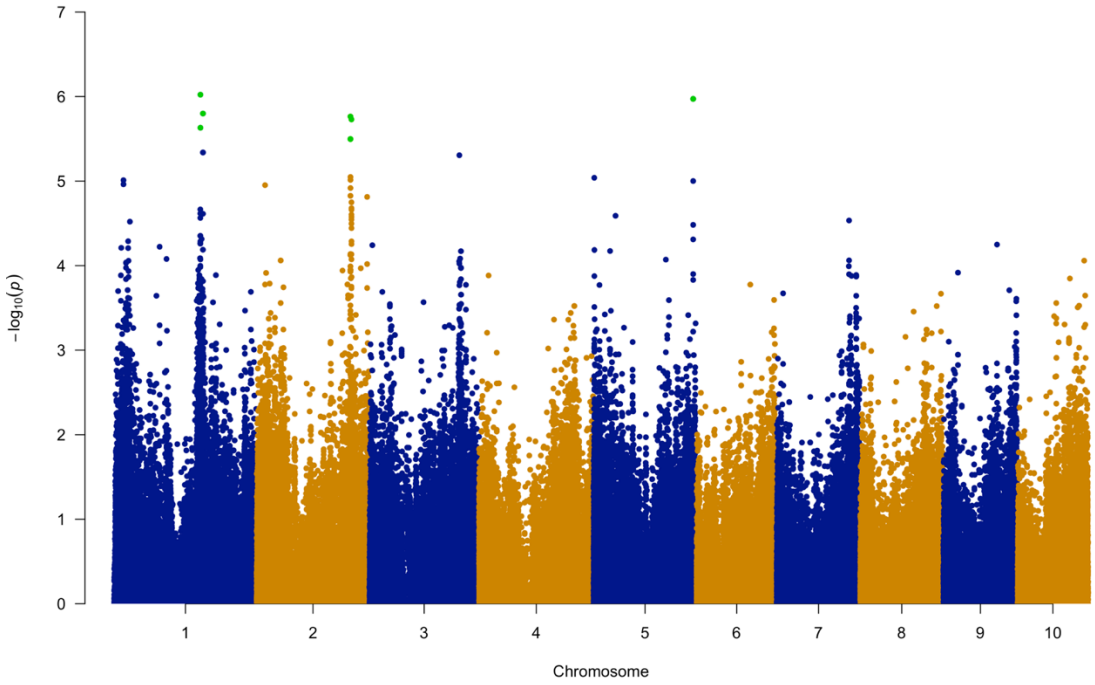

7
6
5
4
3
2
1
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
10
Chromosome

Figure 3.5 Comparison of physical positions of QTL detected in bi-parental linkage mapping conducted by Singh et al. (2016) and significant SNPs from GWAS for chromosomes 1, 2, and 5. The x-axis of each plot represents the physical position, and the y-axis displays the negative $\log_{10}$ of p-value of SNPs in GWAS. Dashed vertical lines indicate the 2-lod support intervals for QTL detected in joint linkage mapping or individual bi-parental family mapping. Significant SNPs in the GWAS are indicated by dark green dots.

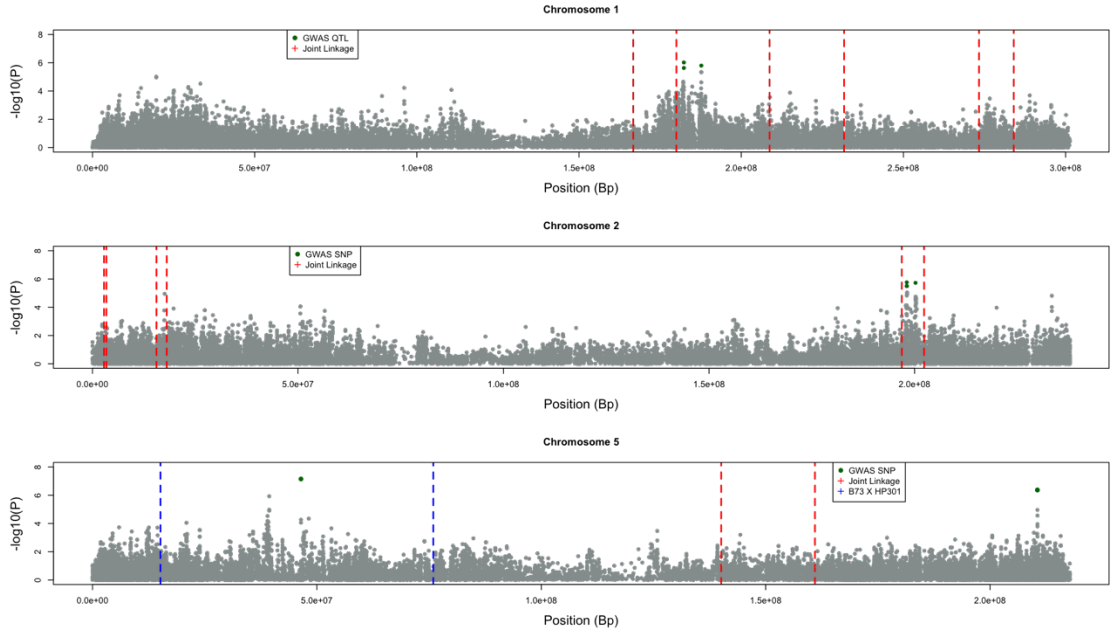
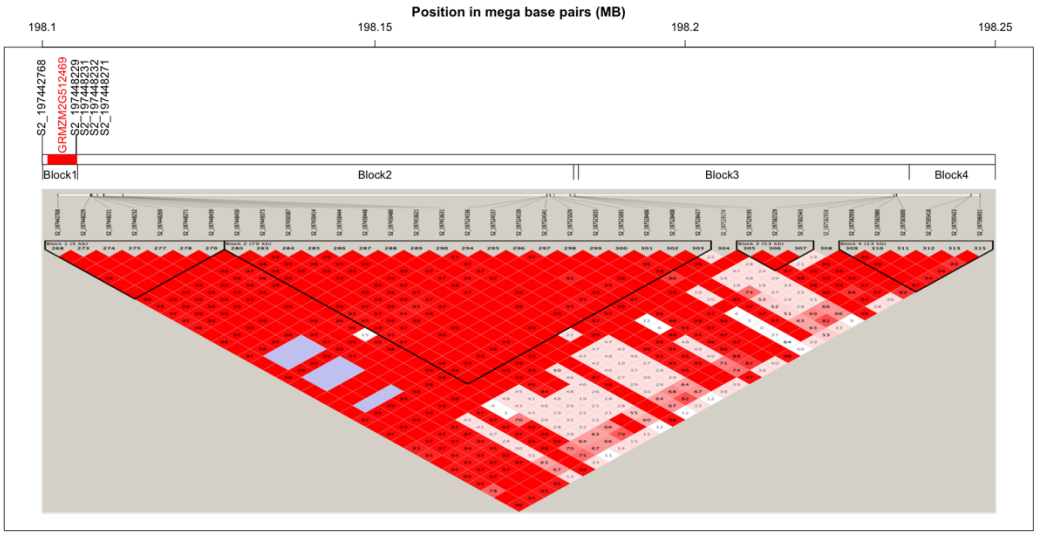


Figure 3.6 Haplotype block analysis of chromosome 2 and 5. (a) Haplotype block analysis of the chromosome 2 region. The region with four significant SNPs consist of four blocks outlined in black. Significant SNPs are displayed in black on the physical map of the region and positions of the candidate genes are displayed in red, (b) haplotype analysis of the chromosome 5 region. This region consist of five haplotype blocks as highlighted in black. Block 2 is the longest with 4 kb.



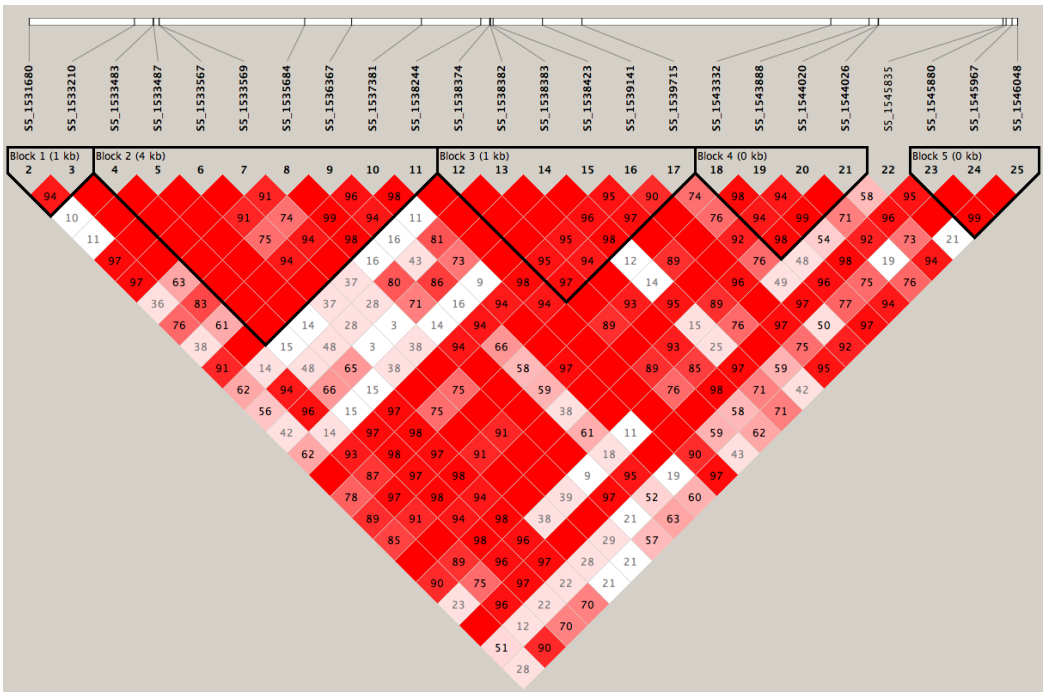


Figure 3.7 Heatmap of eigen genes of 15 modules identified using weighted gene co-expression network analysis (WGCNA). An eigen gene is representative of the gene expression pattern of genes inside that module. For example, module 3 represents the genes that had different expression pattern in the two inbred lines irrespective of the treatment and time point at which a sample was taken.

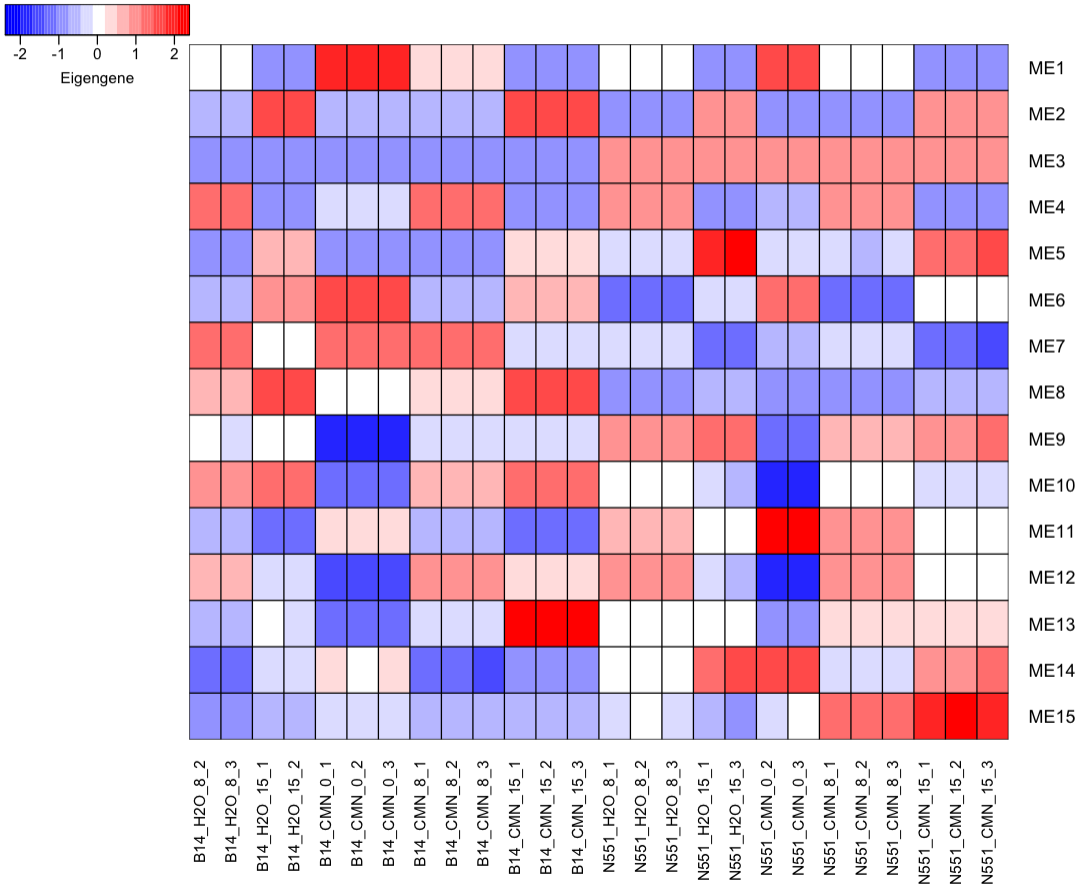
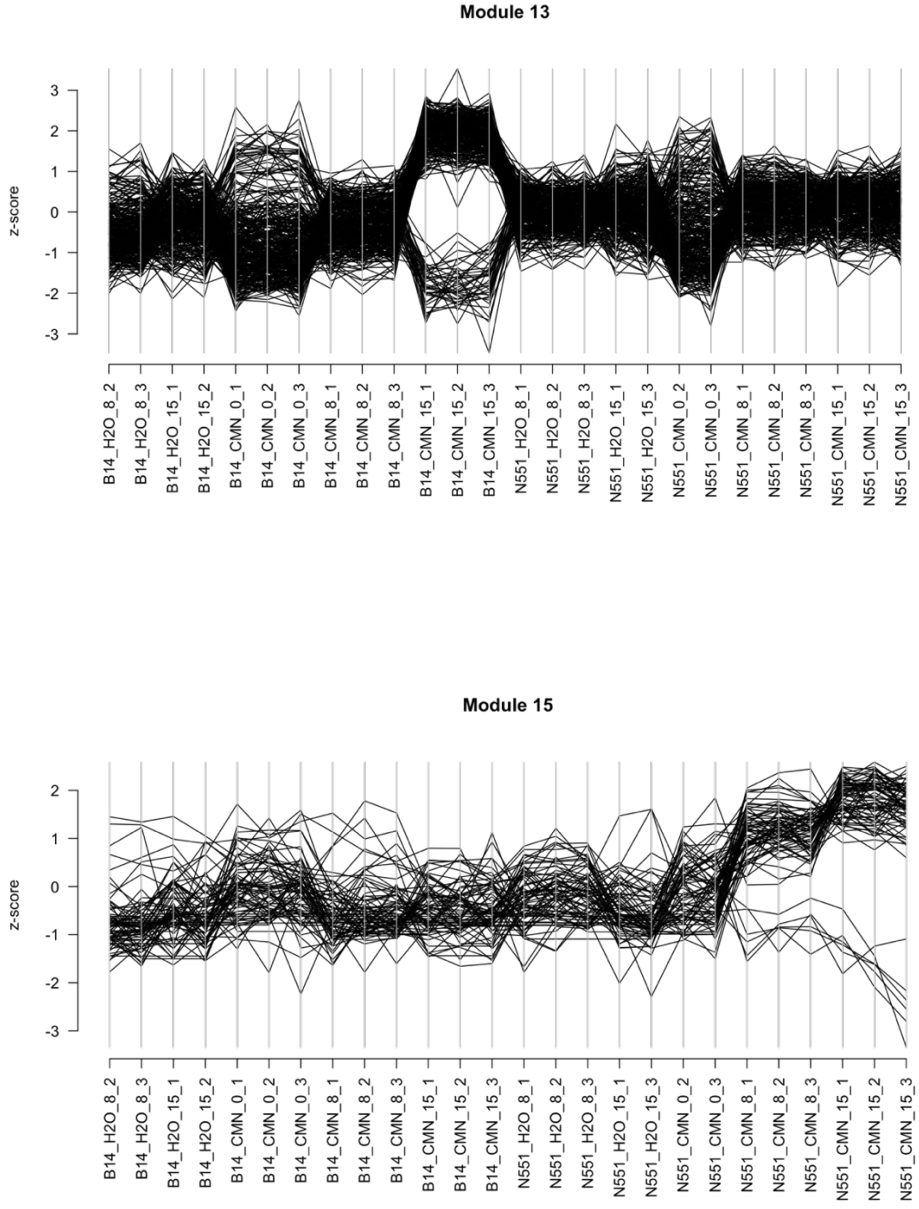


Figure 3.8 Normalized expression of genes inside modules 13 and 15 obtained from WGCNA across all the samples. (a) Normalized expression of genes inside module 13. Large number of genes in this module showed an increased expression at 15 hpi in B14A. A proportion of the genes was also down regulated at 15 hpi in B14A. Genes were not up regulated in N551 in response to *Cmn*, (b) Normalized expression of genes inside module 15. Genes in this module showed differential responses to *Cmn* in inbred line B14A and N551. Expression of most of the genes increases in response to *Cmn* and was highest at 15 hpi in N551.



Chapter Four: Prospects of Genomic Prediction for Resistance to Goss's Bacterial Wilt and Leaf Blight of Maize

4.1 Abstract

Goss's wilt is a bacterial disease of maize caused by a gram positive bacterium *Clavibacter michiganensis* subsp. *nebraskensis* (*Cmn*). Goss's wilt has re-emerged as a disease of economic importance in the western United States and is spreading to other areas of North America. One possible reason for the re-appearance of Goss's wilt could be an increase in susceptibility of germplasm used in commercial maize breeding. Therefore, it is important to screen maize germplasm to identify sources of resistance to Goss's wilt that can be introgressed into commercial maize hybrids, as well as routinely screen and breed for resistance as part of the breeding pipeline. Recently two genome-wide association studies and a linkage mapping study identified several small-effect QTL associated with resistance to Goss's wilt, but no large-effect QTL were identified, indicating the complex genetic architecture of this disease. Genomic prediction, which uses molecular markers uniformly distributed throughout the genome to predict genetic values of selection candidates, may be a good option for improving Goss's wilt resistance. The objective of this study was to assess the prospect of genomic prediction and selection for Goss's wilt resistance in a maize diversity panel of 555 inbred lines and three maize bi-parental populations phenotyped for Goss's wilt and genotyped with genotyping-by-sequencing (GBS). A genomic best linear unbiased prediction model resulted in a predictive ability similar to three other tested models, and hence was used for all subsequent analyses. A predictive ability of 0.56 was achieved in the diversity

panel, and a maximum predictive ability of 0.64 was obtained in B73 x Oh43 population. An analysis of training population composition indicated that much larger training population size is required when predicting across families by adding equal number of lines to the training population. The addition of the training data from the diversity panel minimally impacted the predictive ability within any bi-parental family. Applicability of conditional probabilities that mimicked the plant breeder's ultimate yes/no or keep/discard decision was tested to correctly discard the susceptible lines based on genomic predictions. Probability of discarding susceptible lines was 0.25 and 0.59 in diversity panel and bi-parental dataset. Given that no QTL that explain large proportion of the variation for Goss's wilt was identified using linkage and genome-wide association studies, and that Goss's wilt phenotyping is laborious, genomic selection may be helpful to improve the resistance to Goss's wilt. However, the prospects of genomic prediction and selection need to be studied further by using a large number of bi-parental populations in context of maize breeding programs.

4.2 Introduction

Goss's bacterial wilt and leaf blight of maize, commonly referred to as Goss's wilt has re-emerged as a disease of economic importance in North America (Jackson et al., 2007b). The disease is caused by the gram positive bacterium *Clavibacter michiganensis* subsp. *nebraskensis*. Goss's wilt was first observed in Dawson county in south central Nebraska in 1969 in a maize field and quickly spread within the state as well as to neighboring states, and then later to the rest of Corn Belt by 1981 (Wysong et al., 1973, 1981). The impact of Goss's wilt declined during the 1980s and 1990s, but then

suddenly re-emerged around 2006 and has been since reported in Nebraska, Iowa, Colorado, Missouri, Indiana, Illinois, Kansas, Minnesota, North Dakota, South Dakota, Wisconsin, Texas, Louisiana, Alberta, Manitoba, and Ontario (Ruhl et al., 2009; Malvick et al., 2010; Korus et al., 2011; EPPO, 2014; Friskop et al., 2014a; Howard et al., 2015; Singh et al., 2015; Hosack et al., 2016). Severe yield loss can occur if susceptible hybrids are planted and infection occurs early during the planting season. Up to 44% yield loss has been reported in artificial inoculation experiments with *Cmn* (Carson and Wicks, 1991). Goss's wilt ranked consistently among the top foliage maize diseases during the recent years. Yield loss estimate of 38.5, 103.4, 215.9, and 139.8 million tons have been reported by the Corn Disease Working Group based on statewide disease surveys, university extension feedback, and farmers feedback (Mueller and Wise, 2012, 2013, 2014, 2015).

Speculative reasons for the re-emergence of Goss's wilt include increased acreage of continuous corn, increased adoption of "no till" practices, favorable weather conditions for build-up of inoculum, and increased availability and use of susceptible hybrids (Jackson et al., 2007a; Harveson, 2012). Because the disease only occurred sporadically until 2006, seed companies may have relaxed screening practices and released more susceptible hybrids. According Harveson (2012), there is an association between the re-emergence of Goss's and a reduced availability and utilization of resistant hybrids. In a study conducted to identify possible factors that may have lead to Goss's wilt re-emergence, maize hybrids resistance to Goss's wilt and planting density were

predicted to affect Goss's wilt re-emergence among several other factors evaluated using classification and regression tree analyses (Langemeier et al., 2012)

The use of resistant hybrids is the most effective way to control Goss's wilt (Jackson et al., 2007b). Multiple studies have been conducted to screen maize inbred lines for resistance to Goss's wilt (Schuster et al., 1972a; Calub et al., 1974b). Inheritance of resistance to Goss' wilt appears to be polygenic and additive and thus resistance is easily transferred from inbred parents to hybrids (Martin et al., 1975; Treat et al., 1990; Ngong-Nassah et al., 1992). Breeding techniques that exploit additive gene action and screening over several locations, due to significant genotype-by-environment effects, have been recommended for breeding for resistance to Goss's wilt by Martin et al. (1975).

Unlike other maize diseases, little research using molecular markers to study the genetic basis of Goss' wilt resistance has been conducted because of its lack of importance until recently. Its recent re-emergence, however, has led to two studies that used molecular markers to detect QTL for Goss's wilt resistance. Schaefer and Bernardo, (2013) found nine small effect single nucleotide polymorphism (SNP) markers significantly associated with Goss's wilt in the Minnesota maize inbred panel using a genome-wide association study (GWAS). Eleven quantitative trait loci (QTL), each of small effect, were identified in a joint linkage mapping study conducted by Singh et al., (2016). Each of these eleven QTL explained less than 10% of the phenotypic variation for Goss's wilt (Singh et al., 2016). These results suggest that resistance to Goss's wilt is a complex trait controlled by many loci, and large-effect QTL may not exist, although

this needs to be confirmed through more research on specially designed mapping populations made by crossing highly susceptible and resistant parents.

The combination of GWAS in diverse germplasm and linkage mapping in bi-parental populations is a powerful approach for the dissection of the genetic basis of any quantitative trait. Specific QTL of large effect can be validated and then deployed in breeding programs (Rawat et al., 2016). However, when a trait is controlled by many small-effect QTL, selecting for a particular QTL may not be an effective strategy because small QTL effects are inconsistent and difficult to detect. Moreover, pyramiding QTL together into a single line becomes more difficult as the number of QTL increases (Bernardo, 2010). Genomic selection is an alternative approach, whereby genome-wide molecular markers are used to build a statistical predictive model by combining phenotypic and genotypic data collected on a "training population" (TP). A genomic prediction model is used to predict the genetic value of individuals in a target population, and selections are made on the predictions much like selections are made on phenotypes in a breeding program (Heffner et al., 2009). Key potential advantages of genomic selection over phenotypic selection include reduced cost of evaluation (i.e., genome-wide markers are less expensive than phenotyping, depending on the trait) and reduced time to selection through the use of off-season nurseries combined with genomic predictions of performance.

Although genomic prediction had been initially studied in maize on traits such as grain yield, grain dry matter, biomass, grain moisture, and plant height (Albrecht et al., 2011; Riedelsheimer et al., 2013; Krchov et al., 2015), it could potentially be even more

advantageous for helping to improve disease resistance in maize given that most diseases of maize are quantitative in nature, and disease nurseries and evaluation can be particularly expensive and labor intensive. Genomic prediction has been studied for four maize diseases: northern corn leaf blight (NCLB) (Technow et al., 2013), maize lethal necrosis (MLN) (Gowda et al., 2015), *Fusarium* ear rot (Zila, 2014), and *Gibberella* ear rot (Riedelsheimer et al., 2013). An initial genomic prediction study of *Gibberella* ear rot resistance concluded that within-population prediction was a feasible approach and worked according to the theoretical expectations (Riedelsheimer et al., 2013). Despite higher phenotypic selection accuracy of MLN, it was concluded that routine use of genomic selection (GS) for MLN resistance breeding may be more efficient than phenotypic selection for increasing genetic gain on per year basis (Gowda et al., 2015). A large proportion of phenotypic variation ranging from 31 - 57% was explained by a genomic relationship matrix for *Fusarium* ear rot as compared to only 3 - 12% of the phenotypic variation explained by each of the significant SNPs identified in a GWAS on a diversity panel of 279 maize inbred lines (Zila et al., 2013). Hence, the authors recommended GS as a potential strategy rather than targeting small-effect QTL for *Fusarium* ear rot resistance breeding in maize (Zila, 2014). In GWAS of Goss's wilt, significant SNPs explained only 10% of the phenotypic variation, while a genomic relationship matrix explained 64% of the phenotypic variation. In addition to this complex genetic architecture, laborious phenotyping of Goss's wilt, involving at least three persons for inoculations and extensive lab work for inoculum preparation, prompted this study to explore the prospects of GS for Goss's wilt resistance.

The exact way in which genomic selection for Goss's wilt resistance could be integrated into a maize breeding program would greatly vary by program. A typical number of crosses made during each year in maize breeding programs is not fixed and vary across breeding programs, but the number of crosses is often large such that a large number of breeding families is generated and screened. Bernardo (2010) outlined an example of a doubled haploid (DH)-based commercial maize breeding program in which 80 F_2 or BC_1 populations were used and 4000 DHs were created. While it is desirable in a breeding program to discard the susceptible lines during the early stages, it is not feasible to phenotype all these DH lines for Goss's wilt resistance. Hence, genomic prediction could be helpful to reduce the phenotyping effort and increase the efficiency. Rather than phenotypically screening all DH lines in a disease screening nursery, a subset of the lines could be phenotyped to make a training population and predictions could be made on the remaining lines, from which culling decisions would be made.

The main objective of this study was to assess the genomic prediction accuracy of Goss's wilt resistance within a diversity panel of maize inbred lines as well as three bi-parental populations (Singh et al., 2016). Using the three families, the prospects of discarding susceptible lines based on genomic prediction was explored as it could save a large amount of phenotyping resources for Goss's wilt and increase the efficiency of breeding programs. Finally, the effect of combining lines across families and including a diverse set of lines into a TP on prediction accuracy was assessed.

4.3 Materials and Methods

4.3.1 Plant Materials

The germplasm used in this study is described in detail in a linkage mapping study (Singh et al., 2016) and GWAS (Chapter 3). Data generated from the linkage mapping families and the diversity panel was used in this study. Briefly, the seed for F_5-derived recombinant inbred lines (RILs) from three bi-parental linkage mapping families -- B73 x Oh43, B73 x HP301, and B73 x P39 -- was obtained from the North Central Regional Plant Introduction Station (NCRPIS) located at Ames, IA. B73 is moderately resistant to Goss's wilt, while the other three parents are comparatively susceptible (Singh et al. 2016). A panel of 555 diverse inbred lines which was also used for a GWAS was the other dataset used. The seed for 400 lines was obtained from the NCRPIS and for rest of the 155 lines was provided by Dow AgroSciences.

4.3.2 Experimental Design of Field Evaluations

While B73 x Oh43 bi-parental family was screened for Goss's wilt in three years from 2012 to 2014, the B73 x HP301 and B73 x P39 families were evaluated in two years during 2013 and 2014. In 2012, a Goss's wilt nursery was established at O'Neill, NE by Dupont Pioneer and during rest of the years the experiments were conducted at Mead, NE. Partially replicated completely randomized designs were used for screening of bi-parental families in which the check inbred lines including B73, Oh43, P39, and B14A were replicated multiple times to obtain a better estimate of Goss's wilt infection. The details of the number of recombinant inbred lines used for each family and number of replications for check inbred lines were discussed in Singh et al. (2016).

The panel of diverse inbred lines was evaluated for Goss's wilt in 2014 and 2015. In 2014, 400 lines were planted in a randomized complete block design with three replications at Mead, NE. In 2015, an additional 155 lines were added to the experiment for which the seed was available, to increase power of the panel by increasing the population size. Inbred line B14A and two proprietary Dow AgroSciences inbred lines were used as checks. Detailed description of the association panel selection is provided in the GWAS (Singh, 2017). Briefly, the initial set of 2815 inbred lines was reduced to 900 lines by selecting the lines that flowered within an eight-day window of the flowering of B73. The 900 lines were then divided into 400 clusters using *k*-means clustering and one line from each cluster was selected to achieve the final panel size of 400.

4.3.3 Inoculation and Phenotyping

Mixture of five *Cmn* isolates was used to prepare the inoculum for field inoculations. Isolates collected in Nebraska including 225A, 225B, 225C, 10B, and 194C were plated out on nutrient broth yeast extract agar (NBY) media (Vidaver, 1967). After approximately four days of growth at 25 °C, isolates were tested for virulence in the greenhouse on a susceptible sweet corn hybrid Golden Cross Bantam, planted in 15-cm clay pots filled with steam pasteurized soil. This test was carried out to make sure that the isolates are virulent and capable of producing symptoms before using them for field inoculations. Bacterial colonies from each plate were suspended in 5 mL of 10 mM phosphate buffer at pH of 7.1 (Carlson et al., 1979). Inoculum concentrations were adjusted to 1.0×10^7 CFU/mL using a spectrophotometer at 620 nm wavelength (Schuster, 1975; Suparyono and Pataky, 1989). Bacterial isolates were re-isolated from

the symptomatic plants onto NBY media. Re-isolations were done to ensure that the isolates were virulent before proceeding with field inoculations. A single colony from the re-isolated isolates was used to make new plates in bulk to inoculate the field experiment. Equal amount of each of the five isolates was mixed to prepare the inoculum.

In 2012, inoculations of B73 x Oh43 family at O'Neill was done by DuPont Pioneer using proprietary technology. Following inoculations, one disease rating was collected 54 days after inoculations (DAI) using a rating scale of 1-9 given by Suparyono and Pataky (1989) where 1 indicates a resistant plot and 9 indicates completely dead plot. The scale is described in detail elsewhere (Singh et al., 2016). In 2013, the inoculations were conducted when most of the plants were at V6 growth stage. Motorized weed whippers were used to create wounds on leaves of the plants followed by spraying the inoculum with a backpack sprayer within seconds of wounding the plants. Two disease rating at 15 DAI and 30 DAI were conducted using the same rating scale. Inoculation and disease rating procedure remained same during 2014 and 2015 inoculations and three disease ratings were conducted at 15, 30, and 45 DAI.

4.3.4 Genotype data

A genotypic data set, ZeaGBSv2.7, containing 955,690 SNPs scored using genotyping-by-sequencing was downloaded from panzea.org. This data set included SNP data on each line in the diversity panel as well as the three bi-parental populations used in this study as these were part of the NAM population (Yu et al., 2008). Genotyping by sequencing library preparation and SNP calling was conducted as previously reported (Elshire et al., 2011; Romay et al., 2013; Glaubitz et al., 2014). Missing SNP calls were

imputed using the haplotype-based imputation method Fast Inbred Line Library ImputatioN (FILLIN) (Swarts et al., 2014). The data set still contained missing SNP scores after imputation. Therefore, SNPs were removed if their frequency of missing data exceeded 80% or minor allele frequency was less than 0.05 within the subset of data consisting of the diversity panel and bi-parental populations. The filtered dataset consisted of 342,419 SNPs which was used for all genomic prediction analyses described below.

4.3.5 Phenotypic data analysis

Multiple Goss's wilt ratings within each year and for each dataset (diversity panel and bi-parental populations) were used to calculate the weighted mean disease (WMD). The average of two consecutive ratings was taken, multiplied by the number of days between the two ratings, the values obtained were added, and finally divided by total number of days of evaluation to obtain WMD (Balint-Kurti et al., 2010).

The following model was used to analyze WMD values in both the diversity panel data set and bi-parental population data sets separately:

$$Y_{ijk} = \mu + E_i + \beta_j + r_{k(i)} + \gamma_{ij} + \varepsilon_{ijk}$$

where Y_{ijk} represents the WMD value of the j^{th} inbred grown in the i^{th} environments and k^{th} rep nested in the i^{th} environment; μ represents the intercept; E_i represents the effect of the environment (fixed); β_j represents the effect of the genotype (random), where genotype refers to an accession in the case of the diversity panel and recombinant inbred line in the case of the bi-parental populations; $r_{k(i)}$ represents the effect (fixed) of replication nested within environment; γ_{ij} is the interaction effect (random) of genotypes

with the environment; and ε_{ijk} represents the residual. Best linear unbiased predictions were calculated under the assumption that genotype effects were independent and identically distributed Gaussian random variables. Best linear unbiased predictors calculated using the above model were treated as phenotypes in the genomic prediction models and cross validations.

4.3.6 Genomic prediction models

To choose the best genomic prediction model for optimizing the TP size and composition in bi-parental families, predictive ability of four models was evaluated. The models used were Genomic Best Linear Unbiased Prediction (GBLUP) (VanRaden, 2008), Reproducing Kernel Hilbert Spaces Regression (RKHS) (Gianola et al., 2006), BayesB (Habier et al., 2011), and Neural Network (MacKay, 1992). GBLUP models similarity among the individuals through a realized additive genomic relationship matrix calculated from the genome-wide marker data. GBLUP assumes that marker effects are sampled from a common normal distribution and hence marker effects are assumed to have equal variance. This model is expected to perform well for polygenic traits. BayesB relaxes the assumptions of GBLUP and allows markers to be sampled from different distributions of different variance, including from a point mass distribution with an effect size of zero. RKHS is a semi-parametric model which was proposed by Gianola et al., (2006) to predict quantitative traits from genomic data. RKHS can capture the interaction between the markers without explicitly modeling the interaction effects. RKHS is flexible such that it can use different kernels such as pedigrees, covariates, and molecular data (Gianola and de los Campos, 2008). For example, for predicting genetic value of

individuals in a GBLUP framework, genomic relationship matrix inferred from molecular marker data can be entered as a kernel into RKHS. Neural network (NN) is a non-parametric machine learning method that can model additive and dominance effects at the same time and can be a method of choice for complex traits for which inter loci interaction or epistasis is important (Gianola et al., 2011).

A 10-fold cross validation procedure was used to assess the predictive ability of four different models. In this technique of model validation, the whole dataset was divided into 10 folds each of which was used as a validation set and rest of the lines were used as training set to predict the validation set. Each validation set was predicted five times to calculate the standard deviation of the predictive ability. Pearson correlation coefficient between the average of the genomic estimated breeding values (GEBV) from five replications and the best linear unbiased prediction values was calculated to obtain the predictive ability of the models. Bi-parental families, and diversity panel datasets were used to assess the predictive ability of the models.

All genomic prediction models were implemented using R. GBLUP was implemented using the R package *rrBLUP* (Endelman, 2011); RKHS and BayesB were implemented in the package BGLR (Pérez and de los Camplos, 2013). As RKHS estimates the kernel density function, it requires a positive real value smoothing or bandwidth parameter. The value of the smoothing parameter has an influence on the smoothing of the kernel density estimation and therefore optimum value of smoothing parameter need to be chosen to avoid under or over smoothed estimates. Different values of smoothing parameter (h) ranging from 0.1 to 5 with increments of 0.5 were fit for the

RKHS model and values resulting in highest predictive ability assessed using cross validation were used in the final analysis. For BayesB model, default value of $\pi = 0.5$ was used, which represents the prior proportion of non-zero marker effects. The package brnn was used to fit the neural network model (Pérez-Rodríguez et al., 2013).

4.3.7 Optimizing training population size and composition

The four genomic prediction models tested performed equally well, and hence the GBLUP model was used for all subsequent analyses. The effect of TP size on genomic predictive ability was tested by performing within-family predictions. For each of the bi-parental families, a range of TP size from N=10 to N=100 was used to predict the remaining lines within each family. This was repeated 100 times at each TP size by randomly sampling the lines of the TP without replacement. For example, B73 x Oh43 family has 195 RILs and therefore TP size of 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 was used to predict 185, 175, 165, 155, 145, 135, 125, 115, 105, and 95 lines respectively.

The potential of genomic prediction by combining the lines across families when the number of lines are not enough within family, as may be the case in advanced stages of a typical breeding program, was assessed. Under this scenario, efficiency can be maximized by keeping the TP size small per family and leveraging information from different families. To test this TPs composed of equal number of lines from each family were constructed. A combined-family TP was compared to a single-family TP by measuring predictive ability on a single target family.

Addition of diverse lines to the TP may also be helpful in increasing predictive ability by using the relatedness between the TP and the diverse lines. The lines of diversity panel were combined to each TP size (10-100) for each of the three families to assess the effect on predictive ability. Further, diversity panel was mined to add a specific related sub population to each family. For example, popcorn lines from the diversity panel were combined to TPs for B73 x HP301 family as HP301 is a popcorn line. Similarly, stiff stalk sub population was combined with the B73 x Oh43 family to assess the changes in predictive ability.

4.3.8 Probability of correctly discarding lines based on genomic predictions

The usefulness of genomic prediction models is usually assessed using the predictive ability (accuracy) which is calculated as the Pearson correlation coefficient between the predicted value and the phenotype (true genetic value) of an individual. This metric is ideal when the goal of genomic selection is to rank individuals for quantitative trait values and select a certain proportion with the highest predictive values for advancement. However, a more common use of genomic selection when applied to disease resistance may be to simply identify susceptible lines for culling. In this case, the breeder is interested in the expected proportion of times a correct decision is made versus an incorrect decision. In order to further assess the applicability of genomic prediction to help make this yes/no decision, conditional probabilities were calculated. The probabilities to correctly discard the susceptible lines and correctly retain the resistant lines using a genomic prediction given that the lines were observed to be susceptible and

resistant respectively were estimated. The probability of correctly discarding susceptible lines given that they are truly susceptible is given by:

$$p(S_p|S_t) = \frac{S_p \cap S_t}{S_t}$$

[1]

In [1], S_p is the number of lines predicted to be susceptible and S_t is the number of lines defined as truly susceptible as determined by their phenotypic value. Similarly, the probability of correctly retaining resistant lines given that they were phenotypically categorized as resistant is:

$$p(R_p|R_t) = \frac{R_p \cap R_t}{R_t} \quad [2]$$

In [2], R_p is the number of lines predicted to be resistant using the genomic prediction model and R_t is the number of lines that were defined as resistant based on the phenotypic data. The correlation of the observed phenotypic values and the genomic estimated breeding values was plotted for the combined set of three families, each family separately, and the diversity panel. The observed and predicted susceptibility thresholds were calculated based on the observed and predicted performance of the susceptible check, B14.

4.4 Results

4.4.1 Model Comparison

In the present study, four tested GS models performed equally as observed in other studies. Depending upon the population used, the predictive ability ranged from 0.48 to 0.65 (Table 4.1). GBLUP model gave a predictive ability of 0.57, and 0.65 in diversity panel, and bi-parental populations, respectively, which is either equal to or

greater than the predictive ability obtained using the other three models tested. GBLUP alone was used for the subsequent analyses.

4.4.2 Training Population Size and Composition

As expected, an increase in predictive ability was observed with increasing TP size. Results from the three families indicated that much gain in predictive ability was achieved up to a TP size of 60, and no increase in predictive ability was observed when the TP size was increased further (Figure 4.1). The objective to combine information across the families while keeping the smallest number of lines within family to maximize efficiency was then tested. For each of the three families, the benefit of combining the lines across families was assessed by comparing the predictive ability obtained from the combined TP with equal number of lines from each family to the within family predictive ability obtained with same number of lines in the TP. Minor non-significant gains in predictive ability were observed in each of the three families (Table 4.2). For example, when B73 x Oh43 family was predicted with a combined TP size of 30, a predictive ability of 0.31 was achieved which was similar to the predictive ability of 0.28 obtained for within Oh43 x B73 family at TP size of 10 (Table 4.2).

Together these results indicated that when predicting Goss's wilt resistance, the lines across families can be combined when the number of lines are less within each family as is the case in advanced stages of a typical breeding program. However, much larger combined TP would be required to achieve significant gains in predictive ability. It might be a better to perform within family prediction when sufficient number of lines are available for each family.

4.4.3 Effect of Diverse Lines on Predictive Ability

At TP sizes ranging from 10 to 120, when 555 lines from the diversity panel were included, maximum gain of 0.02 to 0.05 in predictive ability was observed (Table 4.3). For example, adding the 555 lines from the diversity panel to a B73 x Oh43 TP of size 30 only increased the predictive ability from 0.43 to 0.47. Similarly, adding the data from the diversity panel to a B73 x HP301 TP of size 30 only increased the predictive ability from 0.18 to 0.22 (Table 4.3). In the B73 x P39 family, a reduction in predictive ability was observed when data from the diversity panel were added to TPs of size 20 and above (Table 4.3). The gains in predictive ability however were not statistically significant.

The diversity panel was further mined to determine if adding a specific set of lines to the TP would increase predictive ability. Both parents of the B73 x Oh43 family belong to the stiff stalk heterotic group. Therefore, all 100 stiff stalk lines that were part of 555 diverse lines were combined with the TP to predicted B73 x Oh43 family. The stiff stalk lines were added to variously sized B73 x Oh43 TPs (N=10 to 120) to predict the remaining B73 x Oh43 lines. A minimal gain in predictive ability was observed, up to TP size of 50 ranging from 0.01 to 0.06 which was statistically non-significant. (Table 4.3). Similarly, as one of the parents (HP301) of B73 x HP301 family has popcorn background, a subset of 49 popcorn lines from the diversity panel was added to the TP for predicting B73 x HP301 family. No increase in predictive ability was observed in B73 x HP301 family when popcorn lines were added to the TP (Table 4.3).

4.4.4 Using Genomic Prediction to Select Resistant and Discard Susceptible Lines

Using conditional probability, the number of lines correctly predicted to be susceptible given the lines were susceptible based on their observed performance was calculated for the diversity panel dataset, combined dataset of three families, and each family individually. Conditional probabilities were applied to the observed values and predicted values from 10-fold cross validation by using the observed and predicted values of susceptible check B14A as threshold. This approach mimics the plant breeder's decision to keep or discard lines by growing them in a disease screening nursery, ideally in the early stages of selection to minimize the resources spent on phenotyping susceptible lines. However, the number of lines in the early stages of a breeding pipeline is often too large to handle in a disease screening nursery and therefore we explored the possibility of culling lines using genomic prediction.

The probability of correctly discarding the susceptible lines was 0.59 in the combined bi-parental families dataset as compared to 0.25 in case of the diversity panel dataset (Figure 4.2a, 4.2b). As expected, this probability was related to the predictive ability achieved in each dataset. In B73 x Oh43 family, B73 x HP301 family, and B73 x P39 family, the probability of correctly discarding the susceptible lines was 0.56, 0.80, and 0.53, respectively (Figure 4.3a, 4.3b, 4.3c). In B73 x HP301 family, although the probability of correctly discarding the susceptible lines given that they were susceptible $P(S_P|S_T)$ is high, probability of retaining the resistant lines given that they are truly resistant $P(R_P|R_T)$ was low.

Probabilities were also calculated for correctly retaining resistant lines. These probability estimates were 0.98 and 0.84 in the diversity panel and bi-parental datasets, respectively. Although these probability estimates were more promising as compared to the probability estimates of correctly discarding the susceptible lines, for a disease trait such as Goss's wilt it may be desirable to discard the susceptible lines during early stage from the breeding program rather than selecting the resistant lines. Improvements in phenotyping methods and genomic prediction models is therefore required in order to improve the predictive ability and hence the probability of discarding the susceptible lines correctly.

4.5 Discussion

The four genomic prediction models tested in this study (GBLUP, RKHS, BayesB, and NN) all provided similar predictive abilities. Similar results have been commonly reported in previous studies. For example, average predictive ability obtained from 10 GS models compared by (Heslot et al., 2012) ranged from 0.41 to 0.59. Average predictive ability for different traits from several models including RR-BLUP, Bayesian LASSO, BayesCπ, elastic net, weighted Bayesian shrinkage estimation, empirical Bayes, RKHS, neural network, and random forest, ranged from 0.54 to 0.59. Only the support vector machine model gave a little lower predictive ability of 0.41. Even though GS models performed similarly in previous studies, the models were tested again for Goss's wilt because the genetic architecture of this disease has not been extensively studied until recently. To start with, the possibility of one model outperforming the other could not be ruled out based on information from other traits such as grain yield. For example,

Bayesian models could perform better if QTL effects don't fit a single normal distribution. For Goss's wilt, the predictive ability of BayesB and GBLUP models were similar. This indicates that large-effect QTL do not exist in the populations used in this study, or such QTL are very rare and their effects cannot be captured by the SNP markers. This result is in accordance with results from both linkage mapping (Singh et al. 2016) and GWAS (Schaefer and Bernardo, 2013; Singh, 2017).

Training population size is one of the most important factor that determines the predictive ability of the genomic prediction models. Several studies have looked at the optimum TP size for effectively predicting within as well as across bi-parental families. When predicting for different traits including yield index, lignin percentage, glucose release, grain moisture, plant height, ear height, root lodging, stover index, stalk lodging, and grain yield using empirical datasets within a bi-parental family, a TP size of 60-80 lines was recommended by Lorenzana and Bernardo, (2009). In the present study, higher gain in predictive ability were achieved until a TP size of 60-80 depending upon the family used. One of the objective of the present study was to determine the minimum TP size for genomic prediction of Goss's wilt resistance and determine if the significant amount of phenotyping for Goss's wilt can be reduced using genomic prediction. However, results indicated that a fairly large number of lines are still required to be phenotyped for Goss's wilt to get accurate predictions. It would have been valuable and helpful in reducing the number of lines required to be phenotyped for Goss's wilt if a reasonably high predictive ability could have been achieved at a TP size of 20-40. Although, Bernardo (2010) showed that change in standard error of allele frequencies is

maximum when population size is increased from 2 to 50 meaning that increasing the breeding population size beyond 50 to 100 does not provide higher gains, the number and size of families usually vary among different breeding programs. Genomic prediction of Goss's wilt during early stages of breeding programs for within family prediction therefore can still be useful if the family size is large enough.

An important factor influencing the predictive ability is the genetic relationship between the training and target population (Hickey et al., 2014; Lorenz and Smith, 2015), which influences the linkage disequilibrium between the markers and QTL. Comparatively lower predictive ability was achieved in the diversity panel than the bi-parental families as a large number of historical recombination events have broken down the marker-QTL linkages in the diversity panel. The individuals of the bi-parental datasets used in this study shared a common parent; therefore, fewer opportunities existed for recombination and hence the QTL-marker associations are stronger.

The number of individuals that make it to the advanced stages of a breeding program may be small for each family, and therefore combining the individuals across families may be helpful to increase the predictive ability. In this study however it was observed that predictive ability was marginally higher than within family predictions when equal number of lines from the family being predicted were used in combined TP. This lack of benefit of predicting across families is due to decreased genetic relationships between the TP and the VP. This showed that resources should be allocated to within family predictions when enough lines are available within family.

Training population composition is one of the most important factor that determines the predictive ability of genomic prediction and can be controlled by the breeder (Lorenz and Smith, 2015). When distantly related lines were added to the TP in a barley breeding population, a decrease in predictive ability was observed (Lorenz and Smith, 2015). Adding diverse lines to TP in each of the three families in the present study lead to slight changes in predictive ability. In B73 x Oh43 and B73 x HP301 families, a small increase in predictive ability was observed at all TP sizes when diverse lines were added. But in B73 x P39 family an increase in predictive ability was observed at TP size of 10 and a slight reduction thereafter. Adding specific subsets of related lines to the TP e.g stiff stalk lines to B73 x Oh43 family and popcorn lines to B73 x HP301 family also lead to minor increase in predictive ability. Such minor increase in predictive ability does not warrant adding diverse lines to the TP to perform genomic predictions.

As the genetic architecture of most diseases of maize is highly quantitative, genomic prediction studies for disease resistance traits have recommended genomic prediction to breed for disease resistance in maize. For example, Technow et al. (2013) reported prediction accuracy of 0.70 and 0.69 in dent and flint datasets respectively for NCLB. Also, Gowda et al. (2015) achieved a prediction accuracy of 0.41 and 0.56 in two diversity panels. None of the earlier studies looked at the application of genomic prediction for discarding susceptible lines. An alternate method based on conditional probabilities was explored in this study to assess the success of genomic prediction to cull the susceptible lines. Although the cross validated predictive ability of GBLUP model in diversity panel and individual bi-parental populations ranged from 0.46 – 0.69, the

probability of correctly discarding susceptible lines based on genomic prediction was only a little greater than 0.50 (Figures 4.2 and 4.3). In other words, a coin toss would provide nearly as much success in correctly discarding susceptible lines. This result is not as promising as a breeder would like to have a higher confidence in discarding the lines than what was achieved with the genomic prediction models used in this study. Thus, a genomic prediction model may give a reasonably high predictive ability, however its applicability to discard lines should be evaluated carefully before applying it to the breeding programs. On the other hand, this is likely an underestimate of success as the heritabilities of the Goss's wilt resistance evaluations were less than one, and therefore the phenotype for this trait is not a perfect indicator of genetic value.

To conclude, this is the first study that explored the prospects of genomic selection for resistance to Goss's wilt of maize. Given the knowledge about the genetic architecture of Goss's wilt, and that large-effect QTL were not detected for Goss's wilt resistance in the populations used in this study, genomic selection may be a good approach to improve the breeding populations for resistance to Goss's wilt by increasing the frequency of favorable alleles. The conclusions drawn in this study are based on results obtained from a diversity panel and three bi-parental populations consisting of parents with dent corn, popcorn, and sweetcorn genetic backgrounds. More genomic prediction studies should be conducted with a larger number of bi-parental populations that are representative of specific maize breeding programs in order to further evaluate the suitability of genomic selection for Goss's wilt resistance breeding.

Table 4.1 Predictive abilities for the models GBLUP, RKHS, BayesB, and Neural network (NN) when applied to the diversity panel and bi-parental populations. A 10-fold cross validation procedure was used to assess the predictive ability of the models.

Population	Models			
	GBLUP	RKHS	BayesB	NN
Diversity panel	0.57 ± 0.01	0.57 ± 0.006	0.57 ± 0.008	0.48 ± 0.026
Bi-parental	0.65 ± 0.004	0.65 ± 0.002	0.65 ± 0.004	0.63 ± 0.008

Table 4.2 Change in predictive ability when training population was composed from equal number of lines from each of the three families. Training population size was increased from 30 to 300 and each of the three families were predicted. As a comparison, within-family predictive ability at TP size 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110 and 120 for each family is presented.

Validation population	TP Size and Composition	Predictive ability	TP Size and Composition	Predictive ability
Oh43	NA	NA	Oh43 (10)	0.28 ± 0.10
Oh43	NA	NA	Oh43 (20)	0.39 ± 0.08
Oh43	Oh43 (10) + HP301 (10) + P39 (10) = 30	0.31 ± 0.1	Oh43 (30)	0.43 ± 0.08
Oh43	NA	NA	Oh43 (40)	0.49 ± 0.07
Oh43	NA	NA	Oh43 (50)	0.52 ± 0.06
Oh43	Oh43 (20) + HP301 (20) + P39 (20) = 60	0.42 ± 0.08	Oh43 (60)	0.56 ± 0.05
Oh43	NA	NA	Oh43 (70)	0.58 ± 0.05
Oh43	NA	NA	Oh43 (80)	0.60 ± 0.05
Oh43	Oh43 (30) + HP301 (30) + P39 (30) = 90	0.49 ± 0.07	Oh43 (90)	0.61 ± 0.05
Oh43	NA	NA	Oh43 (100)	0.62 ± 0.05
Oh43	NA	NA	Oh43 (110)	0.64 ± 0.05
Oh43	Oh43 (40) + HP301 (40) + P39 (40) = 120	0.53 ± 0.05	Oh43 (120)	0.66 ± 0.05
Oh43	Oh43 (50) + HP301 (50) + P39 (50) = 150	0.56 ± 0.05	NA	NA
Oh43	Oh43 (60) + HP301 (60) + P39 (60) = 180	0.59 ± 0.05	NA	NA
Oh43	Oh43 (70) + HP301 (70) + P39 (70) = 210	0.60 ±0.05	NA	NA
Oh43	Oh43 (80) + HP301 (80) + P39 (80) = 240	0.62 ±0.04	NA	NA
Oh43	Oh43 (90) + HP301 (90) + P39 (90) = 270	0.63 ± 0.04	NA	NA
Oh43	Oh43 (100) + HP301 (100) + P39 (100) = 300	0.64 ± 0.05	NA	NA
HP301	NA	NA	HP301 (10)	0.18 ± 0.10
HP301	NA	NA	HP301 (20)	0.25 ± 0.10
HP301	Oh43 (10) + HP301 (10) + P39 (10) = 30	0.20 ± 0.11	HP301 (30)	0.31 ± 0.09
HP301	NA	NA	HP301 (40)	0.37 ± 0.08
HP301	NA	NA	HP301 (50)	0.39 ± 0.08
HP301	Oh43 (20) + HP301 (20) + P39 (20) = 60	0.30 ± 0.09	HP301 (60)	0.43 ± 0.07
HP301	NA	NA	HP301 (70)	0.45 ± 0.08
HP301	NA	NA	HP301 (80)	0.48 ± 0.08
HP301	Oh43 (30) + HP301 (30) + P39 (30) = 90	0.37 ± 0.09	HP301 (90)	0.47 ± 0.09
HP301	NA	NA	HP301 (100)	0.49 ± 0.10
HP301	NA	NA	HP301 (110)	0.52 ± 0.11
HP301	Oh43 (40) + HP301 (40) + P39 (40) = 120	0.40 ± 0.08	HP301 (120)	0.51 ± 0.15
HP301	Oh43 (50) + HP301 (50) + P39 (50) = 150	0.45 ± 0.07	NA	NA
HP301	Oh43 (60) + HP301 (60) + P39 (60) = 180	0.46 ± 0.07	NA	NA
HP301	Oh43 (70) + HP301 (70) + P39 (70) = 210	0.49 ± 0.06	NA	NA
HP301	Oh43 (80) + HP301 (80) + P39 (80) = 240	0.50 ± 0.08	NA	NA
HP301	Oh43 (90) + HP301 (90) + P39 (90) = 270	0.51 ± 0.09	NA	NA
HP301	Oh43 (100) + HP301 (100) + P39 (100) = 300	0.52 ± 0.1	NA	NA
P39	NA	NA	P39 (10)	0.28 ± 0.10
P39	NA	NA	P39 (20)	0.41 ± 0.09
P39	Oh43 (10) + HP301 (10) + P39 (10) = 30	0.34 ± 0.12	P39 (30)	0.47 ± 0.08
P39	NA	NA	P39 (40)	0.51 ± 0.08
P39	NA	NA	P39 (50)	0.55 ± 0.07

P39	Oh43 (20) + HP301 (20) + P39 (20) = 60	0.46 ± 0.08	P39 (60)	0.59 ± 0.07
P39	NA	NA	P39 (70)	0.60 ± 0.07
P39	NA	NA	P39 (80)	0.63 ± 0.09
P39	Oh43 (30) + HP301 (30) + P39 (30) = 90	0.51 ± 0.07	P39 (90)	0.65 ± 0.09
P39	Oh43 (40) + HP301 (40) + P39 (40) = 120	0.55 ± 0.06	NA	NA
P39	Oh43 (50) + HP301 (50) + P39 (50) = 150	0.57 ± 0.06	NA	NA
P39	Oh43 (60) + HP301 (60) + P39 (60) = 180	0.60 ± 0.05	NA	NA
P39	Oh43 (70) + HP301 (70) + P39 (70) = 210	0.62 ± 0.07	NA	NA
P39	Oh43 (80) + HP301 (80) + P39 (80) = 240	0.64 ± 0.07	NA	NA
P39	Oh43 (90) + HP301 (90) + P39 (90) = 270	0.65 ± 0.1	NA	NA
P39	Oh43 (100) + HP301 (100) + P39 (100) = 300	0.64 ± 0.13	NA	NA

Table 4.3 Change in predictive ability as training population size and composition was changed. (a) Was increased from 10 to 100 in each of the three bi-parental families, (b) 555 diverse lines were added to the TP at each size, (c) subset of the related lines from the diversity panel were added to TP at each size. Stiff stalk lines were added when predicting B73 x Oh43 and popcorn lines were added when predicting B73 x P39 family.

TP Size	TP (a)	Predictive ability	TP (b)	Predictive ability	TP (c)	Predictive ability
10	Oh43	0.28 ± 0.10	Oh43 + GWAS (555)	0.28 ± 0.09	Oh43 + Stiff Stalk (100)	0.34 ± 0.07
20	Oh43	0.39 ± 0.08	Oh43 + GWAS (555)	0.41 ± 0.09	Oh43 + Stiff Stalk (100)	0.42 ± 0.08
30	Oh43	0.43 ± 0.08	Oh43 + GWAS (555)	0.47 ± 0.07	Oh43 + Stiff Stalk (100)	0.47 ± 0.07
40	Oh43	0.49 ± 0.07	Oh43 + GWAS (555)	0.52 ± 0.07	Oh43 + Stiff Stalk (100)	0.53 ± 0.05
50	Oh43	0.52 ± 0.06	Oh43 + GWAS (555)	0.56 ± 0.06	Oh43 + Stiff Stalk (100)	0.55 ± 0.05
60	Oh43	0.56 ± 0.05	Oh43 + GWAS (555)	0.58 ± 0.06	Oh43 + Stiff Stalk (100)	0.57 ± 0.05
70	Oh43	0.58 ± 0.05	Oh43 + GWAS (555)	0.60 ± 0.05	Oh43 + Stiff Stalk (100)	0.60 ± 0.04
80	Oh43	0.60 ± 0.05	Oh43 + GWAS (555)	0.63 ± 0.05	Oh43 + Stiff Stalk (100)	0.60 ± 0.05
90	Oh43	0.61 ± 0.05	Oh43 + GWAS (555)	0.63 ± 0.05	Oh43 + Stiff Stalk (100)	0.62 ± 0.05
100	Oh43	0.62 ± 0.05	Oh43 + GWAS (555)	0.64 ± 0.05	Oh43 + Stiff Stalk (100)	0.63 ± 0.05
110	Oh43	0.64 ± 0.05	Oh43 + GWAS (555)	0.66 ± 0.05	Oh43 + Stiff Stalk (100)	0.65 ± 0.05
120	Oh43	0.66 ± 0.05	Oh43 + GWAS (555)	0.67 ± 0.05	Oh43 + Stiff Stalk (100)	0.66 ± 0.05
10	HP301	0.18 ± 0.10	HP301 + GWAS (555)	0.22 ± 0.08	HP301 + Popcorn (49)	0.18 ± 0.10
20	HP301	0.25 ± 0.10	HP301 + GWAS (555)	0.27 ± 0.10	HP301 + Popcorn (49)	0.26 ± 0.10
30	HP301	0.31 ± 0.09	HP301 + GWAS (555)	0.34 ± 0.09	HP301 + Popcorn (49)	0.33 ± 0.08
40	HP301	0.37 ± 0.08	HP301 + GWAS (555)	0.37 ± 0.08	HP301 + Popcorn (49)	0.39 ± 0.07
50	HP301	0.39 ± 0.08	HP301 + GWAS (555)	0.41 ± 0.08	HP301 + Popcorn (49)	0.40 ± 0.09
60	HP301	0.43 ± 0.07	HP301 + GWAS (555)	0.43 ± 0.07	HP301 + Popcorn (49)	0.44 ± 0.08
70	HP301	0.45 ± 0.08	HP301 + GWAS (555)	0.46 ± 0.08	HP301 + Popcorn (49)	0.47 ± 0.06
80	HP301	0.48 ± 0.08	HP301 + GWAS (555)	0.46 ± 0.08	HP301 + Popcorn (49)	0.48 ± 0.09
90	HP301	0.47 ± 0.09	HP301 + GWAS (555)	0.49 ± 0.10	HP301 + Popcorn (49)	0.51 ± 0.09
100	HP301	0.49 ± 0.10	HP301 + GWAS (555)	0.51 ± 0.10	HP301 + Popcorn (49)	0.51 ± 0.10

110	HP301	0.52 ± 0.11	HP301 + GWAS (555)	0.52 ± 0.15	HP301 + Popcorn (49)	0.53 ± 0.13
120	HP301	0.51 ± 0.15	HP301 + GWAS (555)	0.55 ± 0.14	HP301 + Popcorn (49)	0.50 ± 0.19
10	P39	0.28 ± 0.10	P39 + GWAS (555)	0.33 ± 0.07	NA	NA
20	P39	0.41 ± 0.09	P39 + GWAS (555)	0.39 ± 0.08	NA	NA
30	P39	0.47 ± 0.08	P39 + GWAS (555)	0.44 ± 0.07	NA	NA
40	P39	0.51 ± 0.08	P39 + GWAS (555)	0.48 ± 0.08	NA	NA
50	P39	0.55 ± 0.07	P39 + GWAS (555)	0.50 ± 0.07	NA	NA
60	P39	0.59 ± 0.07	P39 + GWAS (555)	0.51 ± 0.09	NA	NA
70	P39	0.60 ± 0.07	P39 + GWAS (555)	0.56 ± 0.08	NA	NA
80	P39	0.63 ± 0.09	P39 + GWAS (555)	0.57 ± 0.09	NA	NA
90	P39	0.65 ± 0.09	P39 + GWAS (555)	0.60 ± 0.11	NA	NA
100	P39	0.65 ± 0.10	P39 + GWAS (555)	0.62 ± 0.13	NA	NA

Figure 4.1 Predictive ability plotted against training population (TP) size for the B73 x Oh43, B73 x HP301, and B73 x P39 families.

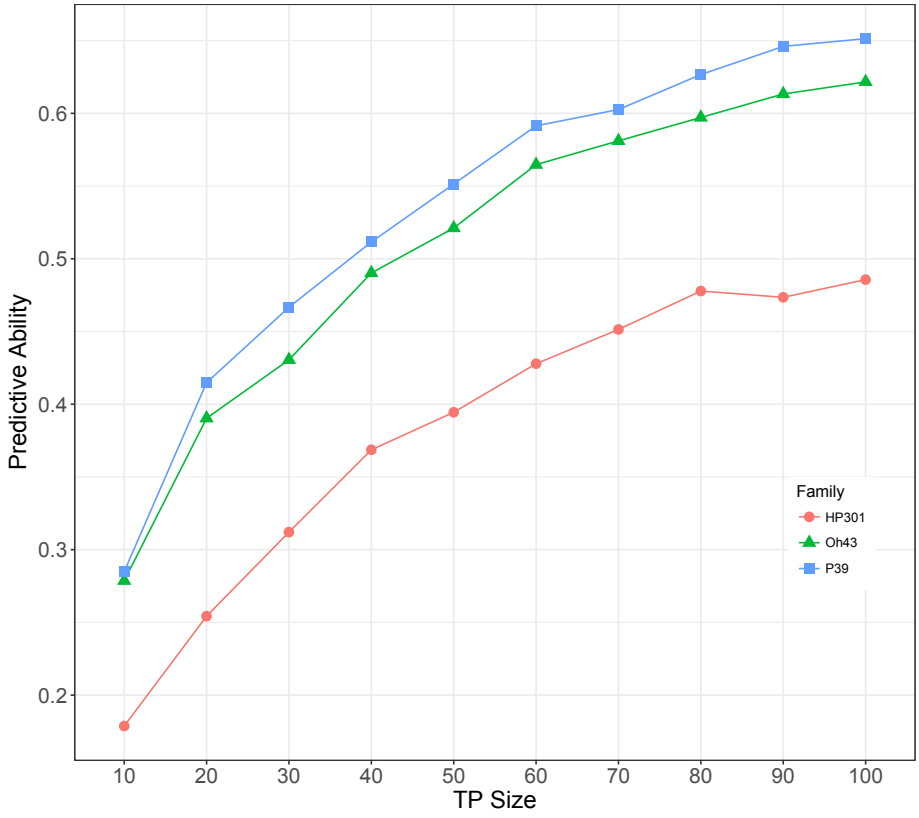
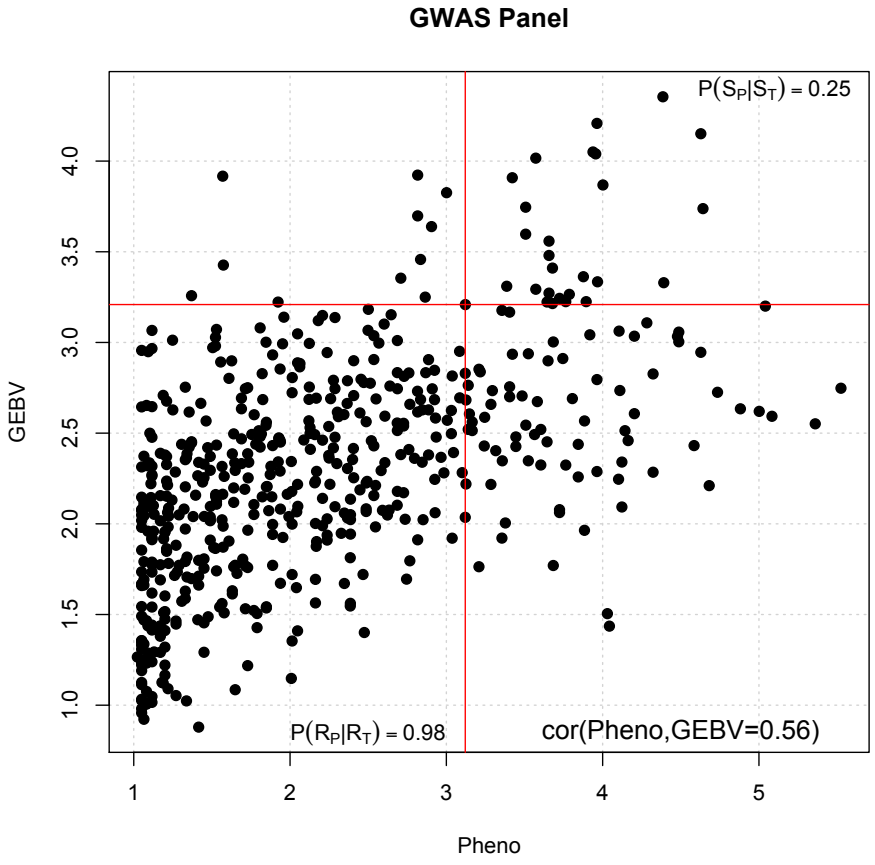


Figure 4.2 Plot of genomic estimated breeding values (GEBV) versus observed phenotypic values. (a) For the diversity panel dataset and (b) three bi-parental families. Threshold lines shown in red were chosen based on the observed and the predicted performance of the susceptible check B14A. $P(S_P|S_T)$ and $P(R_P|R_T)$ are the conditional probabilities that a line is predicted to be susceptible given that it is observed to be susceptible and a line is predicted to be resistant given that it is observed to be resistant, respectively.



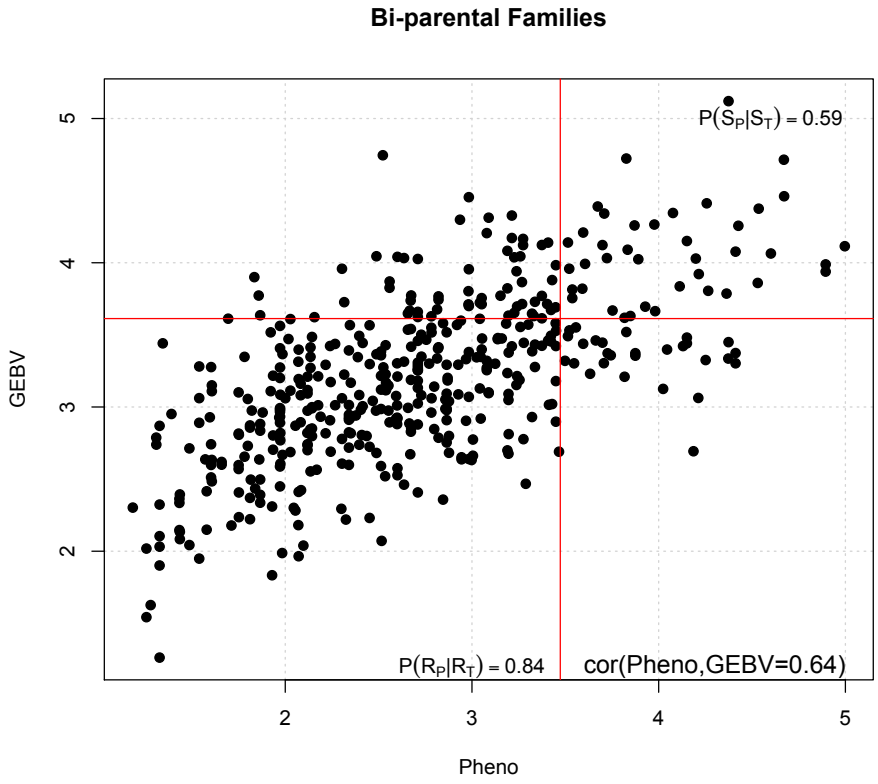
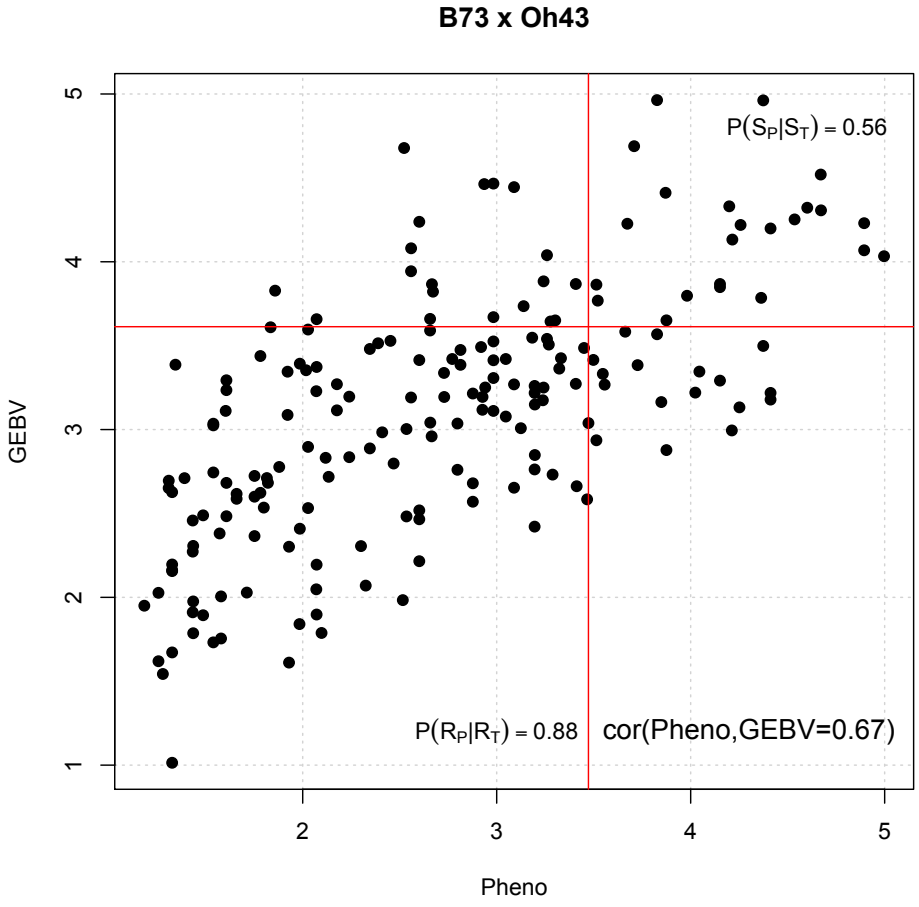
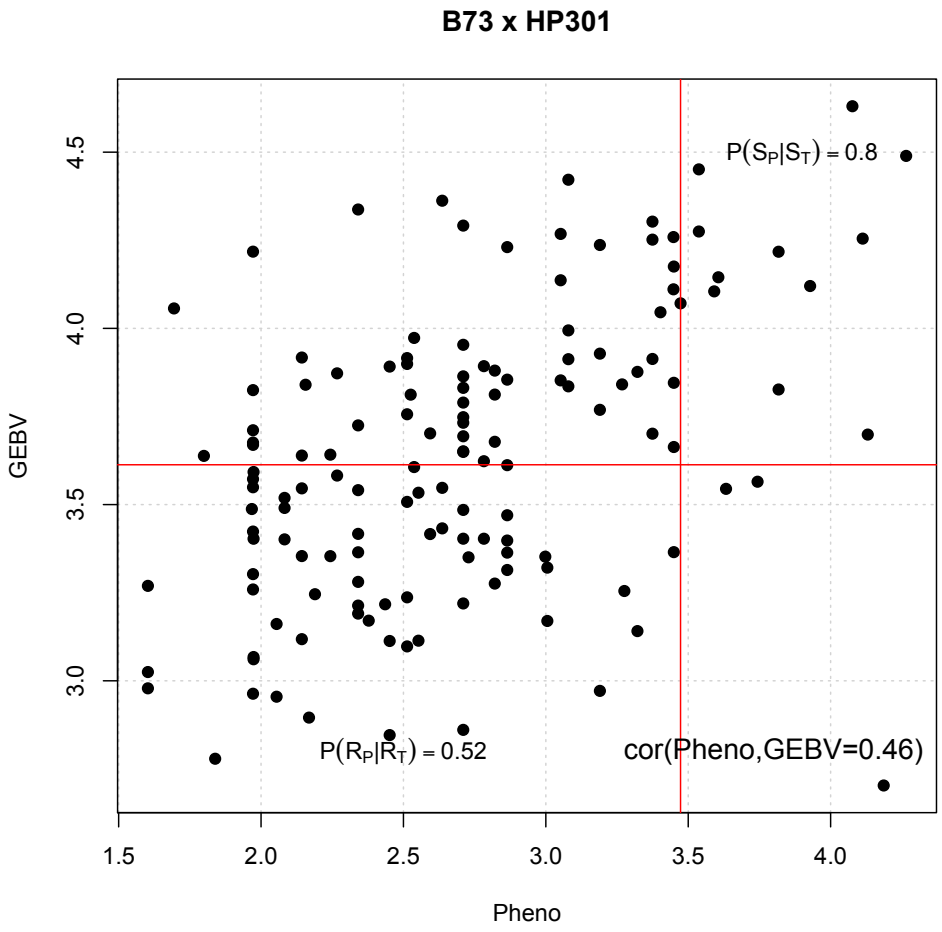
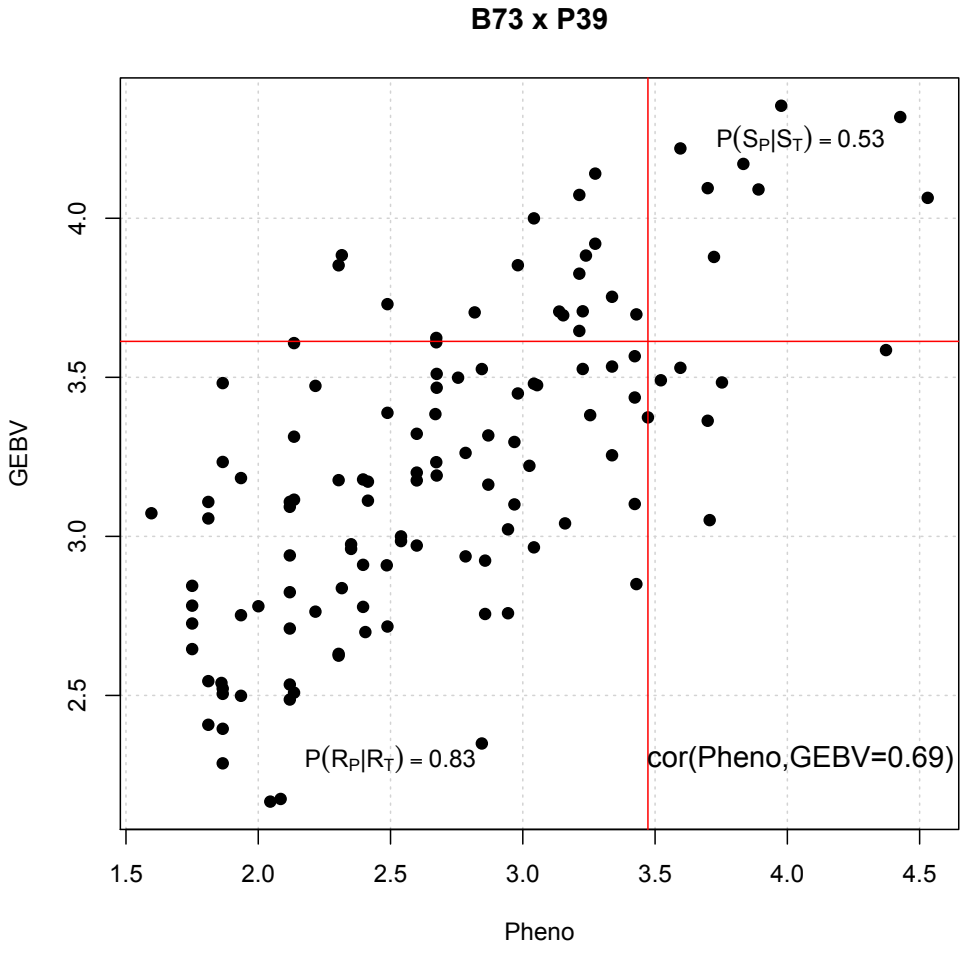

Bi-parental Families
P(S_P|S_T) = 0.59
P(R_P|R_T) = 0.84
cor(Pheno,GEBV=0.64)
GEBV
Pheno
2
3
4
5

Figure 4.3 Plot of genomic estimated breeding values (GEBV) versus observed phenotypic values within each of the three families. (a) for the B73 x Oh43 family, (b) B73 x HP301 family, (c) B73 x P39 family. Threshold lines shown in red were chosen based on the observed and the predicted performance of the susceptible check inbred line B14A. $P(S_P|S_T)$ and $P(R_P|R_T)$ are the conditional probabilities that a line is correctly predicted to be susceptible given that it is observed to be susceptible and a line is correctly predicted to be resistant given that it is observed to be resistant, respectively.




B73 x HP301
GEBV
Pheno
4.5
4.0
3.5
3.0
1.5
2.0
2.5
3.0
3.5
4.0
P(S_P|S_T) = 0.8
P(R_P|R_T) = 0.52
cor(Pheno,GEBV=0.46)


B73 x P39
P(SP|ST) = 0.53
P(RP|RT) = 0.83
cor(Pheno,GEBV=0.69)
GEBV
Pheno
2.5
3.0
3.5
4.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5

Chapter Five: Summary and Final Conclusions

Goss's wilt was first discovered in south central Nebraska in 1969. The disease spread within the Corn Belt states of the United States including Nebraska, Iowa, Kansas, South Dakota, Colorado, Illinois, Minnesota, and Wisconsin by 1981. Varietal resistance was the key that kept the disease under check and prevented it from spreading further. However, as the maize breeding programs quit screening of the maize hybrids for Goss's wilt thinking that it was not a problem anymore, hybrids may have become susceptible overtime during the next 20 years. Several pathogens of maize including *Cmn* are known to survive on surface maize residue (D R Sumner et al., 1981). Therefore, Combination of monocropping of maize and minimum tillage practices may also have contributed to the survival of *Cmn*. Goss's wilt re-emerged as an important disease in the midwestern Corn Belt around 2006 and raised concern among the seed industry. After its resurgence around 2006, Goss's wilt expanded beyond its historical range within the United States Corn Belt. It has been confirmed in 13 states in the United States and three provinces of Canada. Seed industries have been screening their hybrids for Goss's wilt since its reemergence and Goss's wilt resistant products have been introduced to the market for the farmers.

Classical genetic studies using generation mean and diallel analyses have proposed quantitative nature of inheritance of resistance to Goss's wilt. Until recently unlike the other major diseases of maize, genetic studies to understand the genetic basis of resistance to Goss's wilt using molecular markers have been missing likely because the disease was not a concern for commercial maize production. The objectives of this

dissertation were to map QTL using bi-parental mapping populations and a diversity panel, identify differentially expressed genes in resistant and susceptible inbred lines in response to *Cmn*, and asses the prospects of genomic prediction for Goss's wilt.

Joint linkage and linkage mapping identified several small effect QTL associated with resistance to Goss's wilt in three different genetic backgrounds including dent corn, popcorn and sweetcorn. Similarly, three and ten SNPs were associated with resistance to Goss's wilt using the diversity panel alone and the combined dataset of diversity panel and bi-parental families respectively. Each of the SNPs explained less than 10% of the phenotypic variation for Goss's wilt. The results indicated that either a big effect QTL for resistance to Goss's wilt are not present in the populations used in these studies or the frequency of the causal alleles is very low in these populations such that they can not be tagged by the SNPs. Genetic architecture of other major diseases of maize including southern corn leaf blight, northern corn leaf blight, and gray leaf spot have been studied using the NAM population which consist of 25 bi-parental populations with a common parent, B73. These three studies identified QTL with small additive effects which is in accordance with the results obtained from the linkage mapping and GWAS of Goss's wilt (Kump et al., 2011; Poland et al., 2011; Benson et al., 2015).

Identification of two modules of genes with each having 75 and 318 genes respectively using a gene co-expression network analysis that showed differential expression in response to *Cmn* indicated that possibly a large number of genes may be involved in interaction of *Cmn* with maize. This also points towards the complex nature of quantitative disease resistance as hypothesized earlier (Poland et al., 2009).

Any QTL that explains larger proportion of the variation for Goss's wilt was not identified both in linkage and genome-wide association analyses. Such QTL if found, could have been possibly cloned to further dissect the genetic basis of Goss's wilt or deployed in breeding programs after validation. In absence of such QTL in the populations used, genomic prediction and selection can be investigated further for its potential application in maize breeding for resistance to Goss's wilt.

Bibliography

Albrecht, T., V. Wimmer, H.-J. Auinger, M. Erbe, C. Knaak, M. Ouzunova, H. Simianer, and C.-C. Schön. 2011. Genome-based prediction of testcross values in maize. TAG Theor. Appl. Genet. Theor. Angew. Genet. 123(2): 339–350.

Alexander, D.H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res.Available at http://genome.cshlp.org/content/early/2009/07/31/gr.094052.109 (verified 12 June 2016).

Anders, S., P.T. Pyl, and W. Huber. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31(2): 166–169.

Apraku, B.B.-. 1987. A Major Gene for Resistance to Anthracnose Stalk Rot in Maize. Phytopathology 77(6): 957.

Babu, R., S.K. Nair, B.S. Vivek, F.S. Vicente, and B.M. Prasanna. 2012. Integrating marker assisted selection in the DH-based breeding pipeline for rapid development and delivery of superior parental lines and cultivars. In Doubled haploid technology in maize breeding: Theory and practice. Eds. B. M. Prasanna, V Chaikam, and G Mahuku. Mexico D. F. CIMMYT.

Balint-Kurti, P.J., and G.S. Johal. 2009. Maize disease resistance. In Handbook of maize: its biology. Jeff Bennetzen and Sarah Hake (eds.). Springer Science + Business Media, LLC.

Balint-Kurti, P.J., J. Yang, G.V. Esbroeck, J. Jung, and M.E. Smith. 2010. Use of a Maize Advanced Intercross Line for Mapping of QTL for Northern Leaf Blight Resistance and Multiple Disease Resistance. Crop Sci. 50(2): 458–466.

Bao, Y., J.E. Kurle, G. Anderson, and N.D. Young. 2015. Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. Mol. Breed. 35(6): 128.

Barrett, J.C., B. Fry, J. Maller, and M.J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21(2): 263–265.

Benson, J.M., J.A. Poland, B.M. Benson, E.L. Stromberg, and R.J. Nelson. 2015. Resistance to Gray Leaf Spot of Maize: Genetic Architecture and Mechanisms Elucidated through Nested Association Mapping and Near-Isogenic Line Analysis. PLoS Genet. 11(3)Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4357430/ (verified 14 November 2016).

Bentley, S.D., C. Corton, S.E. Brown, A. Barron, L. Clark, J. Doggett, B. Harris, D. Ormond, M.A. Quail, G. May, D. Francis, D. Knudson, J. Parkhill, and C.A. Ishimaru. 2008. Genome of the actinomycete plant pathogen Clavibacter michiganensis subsp sepedonicus suggests recent niche adaptation. J. Bacteriol. 190(6): 2150–2160.

Bernardo, R. 2010. Breeding for quantitative traits in plants. 2nd ed. Stemma Press.

Bernardo, R., and J. Yu. 2007. Prospects for Genomewide Selection for Quantitative Traits in Maize. Crop Sci. 47(3): 1082–1090.

Beyene, Y., K. Semagn, S. Mugo, A. Tarekegne, R. Babu, B. Meisel, P. Sehabiague, D. Makumbi, C. Magorokosho, S. Oikeh, J. Gakunga, M. Vargas, M. Olsen, B.M. Prasanna, M. Banziger, and J. Crossa. 2015. Genetic Gains in Grain Yield Through Genomic Selection in Eight Bi-parental Maize Populations under Drought Stress. Crop Sci. 55(1): 154–163.

Bi, Y.-M., A. Meyer, G.S. Downs, X. Shi, A. El-kereamy, L. Lukens, and S.J. Rothstein. 2014. High throughput RNA sequencing of a hybrid maize and its parents shows different mechanisms responsive to nitrogen limitation. BMC Genomics 15: 77.

Biddle, J.A., D.C. McGee, and E.J. Braun. 1990. Seed transmission of Clavibacter michiganensis subsp. nebraskense in corn. 74: 908–911.

Blanc, G., A. Charcosset, B. Mangin, A. Gallais, and L. Moreau. 2006. Connected populations for detecting quantitative trait loci and testing for epistasis: An application in maize. Theor. Appl. Genet. 113:206–224.

Bonardi, V., and J.L. Dangl. 2012. How complex are intracellular immune receptor signaling complexes? Front. Plant Sci. 3: 237.

Buckler, E.S., J.B. Holland, P.J. Bradbury, C.B. Acharya, P.J. Brown, C. Browne, et al. 2009. The genetic architecture of maize flowering time. Science 325:714–718.

Butler, D., B. Cullis, A. Gilmour, and B. Gogel. 2009. {ASReml}-R reference manual.

Calub, A., W.A. Compton, C.O. Gardner, and M.L. Schuster. 1974a. Reaction of 113 corn inbreds to leaf freckles and wilt of corn. Plant Dis. Report. 58: 956–960.

Calub, A.G., W.A. Compton, C.O. Gardner, and M.L. Schuster. 1974b. Reaction of 113 corn (zea mays) genotypes to leaf freckles and wilt. Plant Dis. Report. 58(10): 956–960.

Carlson, R.R., A.K. Vidaver, D.S. Wysong, and J.H. Riesselman. 1979. A pressure injection device for inoculation of maize with bacterial phytopathogens. Plant Dis. Report. 63: 736–738.

Carson, M.L., and Z.W. Wicks. 1991. Relationship between Leaf Freckles and Wilt Severity and Yield Losses in Closely Related Maize Hybrids. Phytopathology 81(1): 95–98.

Centre for Agriculture and Bioscience International and European and Mediterranean Plant Protection Organization. 2000. *Cla- vibacter michiganensis* subsp. *nebraskensis* [distribution map]. Map No. 549. CAB International, Wallingford, UK.

Chang, R.-Y., and P.A. Peterson. 1995. Genetic Control of Resistance to Bipolaris maydis: One Gene or Two Genes? J. Hered. 86(2): 94–97.

Claflin, L.E., D.L. Bockelman, E.A. Shahin, and T.L. Walter. 1978. The effect of Corynebacterium nebraskense on corn yields. Phytopathol. News 12: 86.

Dekkers, J.C.M., and F. Hospital. 2002. The use of molecular genetics in the improvement of agricultural populations. Nat. Rev. Genet. 3(1): 22–32.

Doerge, R.W., and G.A. Churchill. 1996. Permutation tests for multiple loci affecting a quantitative character. Genetics 142:285–294.

D R Sumner, B Doupnik Jr., and and M.G. Boosalis. 1981. Effects of Reduced Tillage and Multiple Cropping on Plant Diseases. Annu. Rev. Phytopathol. 19(1): 167–187.

Du, Z., X. Zhou, Y. Ling, Z. Zhang, and Z. Su. 2010. agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res. 38(Web Server issue): W64-70.

Eathington, S.R., T.M. Crosbie, M.D. Edwards, R.S. Reiter, and J.K. Bull. 2007. Molecular Markers in a Commercial Breeding Program. Crop Sci. 47(Supplement_3): S-154-S-163.

Eichenlaub, R., and K.-H. Gartemann. 2011. The Clavibacter michiganensis Subspecies: Molecular Investigation of Gram-Positive Bacterial Plant Pathogens. Annu. Rev. Phytopathol. Vol 49 49: 445–464.

Eichenlaub, R., K.-H. Gartemann, and A. Burger. 2007. Clavibacter michiganensis, a group of gram-positive phytopathogenic bacteria. p. 385–421. *In* Gnanamanickam, S.S. (ed.), Plant-Associated Bacteria. Springer Netherlands.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS ONE 6(5): e19379.

Endelman, J.B. 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome 4(3): 250–255.

EPPO. 2014. PQR database. Paris, France: European and Mediterranean Plant Protection Organization. Available at http://www.eppo.int/DATABASES/pqr/pqr.htm.

Forde, B.G., and M.R. Roberts. 2014. Glutamate receptor-like channels in plants: a role as amino acid sensors in plant defence? F1000Prime Rep. 6Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4075314/ (verified 14 February 2017).

Friskop, A., K. Kinzer, M. McConnell, Z. Liu, K. Korus, A. Timmerman, and T. Jackson. 2014a. First Report of Goss's Bacterial Leaf Blight and Wilt of Corn Caused by Clavibacter michiganensis subsp. nebraskensis in North Dakota. Plant Dis. 98(12): 1739–1739.

Friskop, A., K. Kinzer, M. McConnell, Z. Liu, K. Korus, A. Timmerman, and T. Jackson. 2014b. First Report of Goss's Bacterial Leaf Blight and Wilt of Corn Caused by *Clavibacter michiganensis* subsp. *nebraskensis* in North Dakota. Plant Dis. 98(12): 1739–1739.

Gardner, C.O., and S.A. Eberhart. 1966. Analysis and Interpretation of the Variety Cross Diallel and Related Populations. Biometrics 22(3): 439–452.

Gardner, C.O., and M.L. Schuster. 1974. Genetic studies of susceptibility to bacterial leaf freckles and wilt, Corynebacterium nebraskense. Maize Genet. Coop. Newsl. 47: 155–157.

Gianola, D., and G. de los Campos. 2008. Inferring genetic values for quantitative traits non-parametrically. Genet. Res. 90(6): 525–540.

Gianola, D., R.L. Fernando, and A. Stella. 2006. Genomic-Assisted Prediction of Genetic Value With Semiparametric Procedures. Genetics 173(3): 1761–1776.

Gianola, D., H. Okut, K.A. Weigel, and G.J. Rosa. 2011. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. BMC Genet. 12: 87.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PloS One 9(2): e90346.

Gowda, M., B. Das, D. Makumbi, R. Babu, K. Semagn, G. Mahuku, M.S. Olsen, J.M. Bright, Y. Beyene, and B.M. Prasanna. 2015. Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. TAG Theor. Appl. Genet. Theor. Angew. Genet. 128(10): 1957–1968.

Habier, D., R.L. Fernando, K. Kizilkaya, and D.J. Garrick. 2011. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics 12: 186.

Harveson, R.M. 2012. The curious re-emergence of Goss's wilt of corn and bacterial wilt of dry beans in the central plains. Available at http://plantpathology.unl.edu/curious-re-emergence-goss'-wilt-corn-and-bacterial-wilt-dry-beans-central-high-plains.

Heslot, N., H.-P. Yang, M.E. Sorrells, and J.-L. Jannink. 2012. Genomic Selection in Plant Breeding: A Comparison of Models. Crop Sci. 52(1): 146–160.

Hickey, J.M., S. Dreisigacker, J. Crossa, S. Hearne, R. Babu, B.M. Prasanna, M. Grondona, A. Zambelli, V.S. Windhausen, K. Mathews, and G. Gorjanc. 2014. Evaluation of Genomic Selection Training Population Designs and Genotyping Strategies in Plant Breeding Programs Using Simulation. Crop Sci. 54(4): 1476–1488.

Holland, J.B., W.E. Nyquist, and C.T. Cervantes-Martinez. 2003. Estimating and interpreting heritability for plant breeding: An update. Plant Breed. Rev. 22:9–112.

Holland, J.B. 2007. Genetic architecture of complex traits in plants. Curr. Opin. Plant Biol. 10(2): 156–161.

Hosack, P.K., L.E. Sweets, G.L. Miller, M.J. Calcutt, and B.E. Arenz. 2016. First Report of Goss's Bacterial Wilt and Leaf Blight on Corn Caused by Clavibacter michiganensis subsp. nebraskensis in Missouri. Plant Dis.Available at http://dx.doi.org/10.1094/PDIS-02-16-0227-PDN.

Howard, R.J., M.W. Harding, J. Lynn, L.M. Kawchuk, and N.M. Rasmussen. 2015. First Report of Goss's Bacterial Wilt and Leaf Blight on Corn Caused by Clavibacter michiganensis subsp. nebraskensis in Alberta, Canada. Plant Dis. 99(7): 1034.

Jackson, T.A., R.M. Harveson, and A.K. Vidaver. 2007a. Goss's bacterial wilt and leaf blight of corn. NebGuide Coop Ext Serv IANR Univ Nebr Linc. G1675.

Jackson, T.A., R.M. Harveson, and A.K. Vidaver. 2007b. Reemergence of Goss's Wilt and Blight of Corn to the Central High Plains. Plant Health Prog. doi:10.1094/PHP-2007-0919-01-BR.

Jackson, T.A., and J. Rees. 2010. Corn diseases update: Southern rust confirmed and Goss's wilt continues. CropWatch Newsletter August: 6–13.

Jakobsson, M., and N.A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14): 1801–1806.

Jones, J.D.G., and J.L. Dangl. 2006. The plant immune system. Nature 444(7117): 323–329.

Kakumanu, A., M.M.R. Ambavaram, C. Klumas, A. Krishnan, U. Batlang, E. Myers, R. Grene, and A. Pereira. 2012. Effects of Drought on Gene Expression in Maize Reproductive and Leaf Meristem Tissue Revealed by RNA-Seq. Plant Physiol. 160(2): 846–867.

Kang, S., H.B. Kim, H. Lee, J.Y. Choi, S. Heu, C.J. Oh, S.I. Kwon, and C.S. An. 2006. Overexpression in Arabidopsis of a plasma membrane-targeting glutamate receptor from small radish increases glutamate-mediated Ca2+ influx and delays fungal infection. Mol. Cells 21(3): 418–427.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S.L. Salzberg. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14: R36.

Korus, K.A., A.D. Timmerman, R.D. French-Monar, and T.A. Jackson. 2011. First Report of Goss's Bacterial Wilt and Leaf Blight (Clavibacter michiganensis subsp nebraskensis) of Corn in Texas. Plant Dis. 95(1): 73–73.

Krchov, L.-M., G.A. Gordillo, and R. Bernardo. 2015. Multienvironment Validation of the Effectiveness of Phenotypic and Genomewide Selection within Biparental Maize Populations. Crop Sci. 55(3): 1068–1075.

Kump, K.L., P.J. Bradbury, R.J. Wisser, E.S. Buckler, A.R. Belcher, M.A. Oropeza-Rosas, J.C. Zwonitzer, S. Kresovich, M.D. McMullen, D. Ware, P.J. Balint-Kurti, and J.B. Holland. 2011. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat. Genet. 43(2): 163-U120.

Langemeier, C.B. 2012. Improved understanding of factors influencing the re-emergence of Goss's bacterial wilt and blight of corn. MS Thesis Retrieved Theses Diss. Stud. Res. Agron. Hortic. Available Httpdigitalcommonsunleduagronhortdiss58.

Langemeier, C.B., A.E. Robertson, D. Wang, and T.A. Jackson-Zeims. 2012. Factors affecting the development and severity of Clavibacter michiganensis subsp. nebraskensis of Zea mays. Phytopathol. Suppl 5 S5.6: 102.

Langemeier, C.B., A.E. Robertson, D. Wang, T.A. Jackson-Ziems, and G.R. Kruger. 2017. Factors Affecting the Development and Severity of Goss's Bacterial Wilt and Leaf Blight of Corn, Caused by Clavibacter michiganensis subsp. nebraskensis. Plant Dis. PDIS-01-15-0038-RE.

Langfelder, P., and S. Horvath. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9: 559.

Lanubile, A., A. Ferrarini, V. Maschietto, M. Delledonne, A. Marocco, and D. Bellin. 2014. Functional genomic analysis of constitutive and inducible defense responses to Fusarium verticillioides infection in maize genotypes with contrasting ear rot resistance. BMC Genomics 15: 710.

Li, F., J. Wang, C. Ma, Y. Zhao, Y. Wang, A. Hasi, and Z. Qi. 2013. Glutamate receptor-like channel3.3 is involved in mediating glutathione-triggered cytosolic calcium transients, transcriptional changes, and innate immunity responses in Arabidopsis. Plant Physiol. 162(3): 1497–1509.

Lippert, C., J. Listgarten, Y. Liu, C.M. Kadie, R.I. Davidson, and D. Heckerman. 2011. FaST linear mixed models for genome-wide association studies. Nat. Methods 8(10): 833–835.

Liu, Y., Y. Guo, C. Ma, D. Zhang, C. Wang, Q. Yang, and M. Xu. 2016. Transcriptome analysis of maize resistance to Fusarium graminearum. BMC Genomics 17Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4924250/ (verified 16 January 2017).

Lorenz, A.J., and K.P. Smith. 2015. Adding Genetically Distant Individuals to Training Populations Reduces Genomic Prediction Accuracy in Barley. Crop Sci. 55(6): 2657–2667.

Lorenz, A.J., K.P. Smith, and J.-L. Jannink. 2012. Potential and Optimization of Genomic Selection for Fusarium Head Blight Resistance in Six-Row Barley. Crop Sci. 52(4): 1609–1621.

Lorenzana, R.E., and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. TAG Theor. Appl. Genet. Theor. Angew. Genet. 120(1): 151–161.

Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550.

MacKay, D.J.C. 1992. Bayesia Interpolation. Neural Comput. 4: 415–447.

Mahuku, G., B.E. Lockhart, B. Wanjala, M.W. Jones, J.N. Kimunye, L.R. Stewart, B.J. Cassone, S. Sevgan, J.O. Nyasani, E. Kusia, P.L. Kumar, C.L. Niblett, A. Kiggundu, G. Asea, H.R. Pappu, A. Wangai, B.M. Prasanna, and M.G. Redinbaugh. 2015. Maize Lethal Necrosis (MLN), an Emerging Threat to Maize-Based Food Security in Sub-Saharan Africa. Phytopathology 105(7): 956–965.

Makarevitch, I., A.J. Waters, P.T. West, M. Stitzer, C.N. Hirsch, J. Ross-Ibarra, and N.M. Springer. 2015. Transposable Elements Contribute to Activation of Maize Genes in Response to Abiotic Stress. PLOS Genet. 11(1): e1004915.

Mallowa, S.O., G.Y. Mbofung, S.K. Eggenberger, R.L. Den Adel, S.R. Scheiding, and A.E. Robertson. 2015. Infection of Maize by Clavibacter michiganensis subsp. nebraskensis Does Not Require Severe Wounding. Plant Dis. 100(4): 724–731.

Malvick, D., R. Syverson, D. Mollov, and C.A. Ishimaru. 2010. Goss's Bacterial Blight and Wilt of Corn Caused by Clavibacter michiganensis subsp nebraskensis Occurs in Minnesota. Plant Dis. 94(8): 1064–1064.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17(1): 10–12.

Martin, P.R., C.O. Gardner, A.G. Calub, and M.L. Schuster. 1975. Inheritance ofsusceptibility and tolerance to leaf freckles and wilt (Corynebacterium nebraskense) of corn. Maize Genet. Coop. Newsl. 49: 137–138.

Mather, K., and J.L. Jinks. 1982. Biometrical genetics: The study of continuous variation. Cornell University Press.

Matsumura, H., S. Reich, A. Ito, H. Saitoh, S. Kamoun, P. Winter, G. Kahl, M. Reuter, D.H. Krüger, and R. Terauchi. 2003. Gene expression analysis of plant host–pathogen interactions by SuperSAGE. Proc. Natl. Acad. Sci. 100(26): 15718–15723.

McMullen, M.D., S. Kresovich, H.S. Villeda, P. Bradbury, H. Li, Q. Sun, S. Flint-Garcia, J. Thornsberry, C. Acharya, C. Bottoms, P. Brown, C. Browne, M. Eller, K. Guill, C. Harjes, D. Kroon, N. Lepak, S.E. Mitchell, B. Peterson, G. Pressoir, S. Romero, M.O. Rosas, S. Salvo, H. Yates, M. Hanson, E. Jones, S. Smith, J.C. Glaubitz, M. Goodman, D. Ware, J.B. Holland, and E.S. Buckler. 2009. Genetic Properties of the Maize Nested Association Mapping Population. Science 325(5941): 737–740.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157(4): 1819–1829.

Mueller, D., K. Wise, A. Sisson, T. Allen, G. Bergstrom, D. Bosley, C. Bradley, K. Broders, E. Byamukama, D.K. Malvick, M. Chilvers, M. Collins, T. Faske, A. Friskop, R. Heiniger, C. Hollier, D. Hooker, T. Isakeit, T. Jackson-Zeims, D. Jardine, K. Kinzer, M. McMullen, R. Meyer, P. Paul, A. Robertson, G. Roth, D. Smith, C. Tande, A. Tenuta, P. Vincelli, and F. Warner. 2016. Corn yield loss estimates due to diseases in the United States and Ontario, Canada from 2012 to 2015. Plant Health Prog.: 211–222.

Negeri, A.T., N.D. Coles, J.B. Holland, and P.J. Balint-Kurti. 2011. Mapping QTL controlling southern leaf blight resistance by joint analysis of three related recombinant inbred line populations. Crop Sci. 51:1571–1579.

Ngong-Nassah, E.N., M.L. Carson, and Z.W. Wicks. 1992. Inheritance of Resistance to Leaf Freckles and Wilt Caused by Clavibacter-Michiganense Subsp Nebraskense in Early Maturing Maize Inbred Lines. Phytopathology 82(2): 142–146.

Oerke, E.-C. 2006. Crop losses to pests. J. Agric. Sci. 144(1): 31–43.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20(2): 289–290.

Parlevliet, J.E. 1981. Disease resistance and its consequences for plant breeding. In: K.J. Frey (ed.) Plant Breeding II. Iowa State University Press, Ames.

Parlevliet, J.E. 2002. Durability of resistance against fungal, bacterial and viral pathogens; present situation. Euphytica 124(2): 147–156.

Pataky, J.K. 1988. Classification of Sweet Corn Hybrid Reactions to Common Rust, Northern Leaf Blight, Stewart's Wilt, and Goss' Wilt and Associated Yield Reductions. Phytopathology 78(2): 172.

Pérez, P., and G. de los Camplos. 2013. BGLR: A Statistical Package for Whole Genome Regression and Prediction. R Package: 1–30.

Pérez-Rodríguez, P., D. Gianola, K.A. Weigel, G.J.M. Rosa, and J. Crossa. 2013. Technical Note: An R package for fitting Bayesian regularized neural networks with applications in animal breeding. J. Anim. Sci. 91(8): 3522–3531.

Poland, J.A., P.J. Balint-Kurti, R.J. Wisser, R.C. Pratt, and R.J. Nelson. 2009. Shades of gray: the world of quantitative disease resistance. Trends Plant Sci. 14(1): 21–29.

Poland, J.A., P.J. Bradbury, E.S. Buckler, and R.J. Nelson. 2011. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc. Natl. Acad. Sci. U. S. A. 108(17): 6893–6898.

Poland, J., and J. Rutkoski. 2016. Advances and Challenges in Genomic Selection for Disease Resistance. Annu. Rev. Phytopathol. 54(1): 79–98.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, M.J. Daly, and P.C. Sham. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81(3): 559–575.

Ramakrishna, W., J. Emberton, M. Ogden, P. SanMiguel, and J.L. Bennetzen. 2002. Structural Analysis of the Maize Rp1 Complex Reveals Numerous Sites and Unexpected Mechanisms of Local Rearrangement. Plant Cell 14(12): 3213–3223.

Rawat, N., M.O. Pumphrey, S. Liu, X. Zhang, V.K. Tiwari, K. Ando, H.N. Trick, W.W. Bockus, E. Akhunov, J.A. Anderson, and B.S. Gill. 2016. Wheat Fhb1 encodes a chimeric lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to Fusarium head blight. Nat. Genet. 48(12): 1576–1580.

Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, R. Sulpice, T. Altmann, M. Stitt, L. Willmitzer, and A.E. Melchinger. 2012. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat. Genet. 44(2): 217–220.

Riedelsheimer, C., J.B. Endelman, M. Stange, M.E. Sorrells, J.-L. Jannink, and A.E. Melchinger. 2013. Genomic predictability of interconnected biparental maize populations. Genetics 194(2): 493–503.

Robertson, A.E. 2012. Goss's wilt: A 2012 recap and looking ahead to 2013. Proceedings of the 2013 Wisconsin Crop Management Conference 52, Madison, WI. 15–17 Jan. 2013. Soil Science Extension, Univ. Wisconsin-Madison, Madison WI. p. 172–174.

Rocheford, T.R., C.A. Gardner, and A.K. Vidaver. 1985. Chromosomal location of additional genes for resistance to Corynebacterium (Clavibacter) michiganense ssp. nebraskense. Maize Genet. Coop. Newsl. 60: 73–74.

Rocheford, T.R., C.O. Gardner, and A.K. Vidaver. 1989. Genetic studies of resistance inmaize (Zea mays L.) to Goss's bacterial wilt and blight (Clavibacter michiganensis subsp. nebraskensis). J Hered 80: 351–356.

Romay, M.C., M.J. Millard, J.C. Glaubitz, J.A. Peiffer, K.L. Swarts, T.M. Casstevens, R.J. Elshire, C.B. Acharya, S.E. Mitchell, S.A. Flint-Garcia, M.D. McMullen, J.B. Holland, E.S. Buckler, and C.A. Gardner. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 14(6): R55.

Ruhl, G., K. Wise, T. Creswell, A. Leonberger, and C. Speers. 2009. First Report of Goss's Bacterial Wilt and Leaf Blight on Corn Caused by Clavibacter michiganensis subsp nebraskensis in Indiana. Plant Dis. 93(8): 841–841.

Rutkoski, J., J. Benson, Y. Jia, G. Brown-Guedira, J.-L. Jannink, and M. Sorrells. 2012. Evaluation of Genomic Prediction Methods for Fusarium Head Blight Resistance in Wheat. Plant Genome 5(2): 51–61.

Rutkoski, J., R.P. Singh, J. Huerta-Espino, S. Bhavani, J. Poland, J.L. Jannink, and M.E. Sorrells. 2015. Genetic Gain from Phenotypic and Genomic Selection for Quantitative Resistance to Stem Rust of Wheat. Plant Genome 8(2)Available at https://dl.sciencesocieties.org/publications/tpg/articles/8/2/plantgenome2014.10.0074 (verified 10 January 2017).

Sahu, S.S., T. Weirick, and R. Kaundal. 2014. Predicting genome-scale Arabidopsis-Pseudomonas syringae interactome using domain and interolog-based approaches. BMC Bioinformatics 15(11): S13.

Sato, M., K. Tsuda, L. Wang, J. Coller, Y. Watanabe, J. Glazebrook, and F. Katagiri. 2010. Network Modeling Reveals Prevalent Negative Regulatory Relationships between Signaling Sectors in Arabidopsis Immune Signaling. PLOS Pathog. 6(7): e1001011.

Schaefer, C.M., and R. Bernardo. 2013. Genomewide Association Mapping of Flowering Time, Kernel Composition, and Disease Resistance in Historical Minnesota Maize Inbreds. Crop Sci. 53(6): 2518–2529.

Schuster, M.L. 1975. Leaf Freckles and Wilt of Corn Incited by Corynebacterium nebraskense Schuster, Hoff, Mandel, Lazar 1972. Institute of Agriculture and Natural Resources. UNL Res. Bull. 270.

Schuster, M.L., W.A. Compton, and B. Hoff. 1972a. Reaction of corn inbred lines to the Nebraska leaf freckles and wilt bacterium. Plant Dis. Report. 56: 863–865.

Schuster, M.L., B. Hoff, M. Mandel, and I. Lazar. 1972b. Leaf freckles and wilt, a new corn disease. 27th Ann Corn Sorghum Res. Congr Chic. Ill Dec12-14: 176–191.

Segonzac, C., A.P. Macho, M. Sanmartín, V. Ntoukakis, J.J. Sánchez-Serrano, and C. Zipfel. 2014. Negative control of BAK1 by protein phosphatase 2A during plant innate immunity. EMBO J. 33(18): 2069–2079.

Singh, A., A.P. Andersen, T.A. Jackson-Zeims, and A.J. Lorenz. 2016. Mapping quantitative trait loci for resistance to Goss's bacterial wilt and leaf blight in North American Maize by joint linkage analysis. Crop Sci. 56.

Singh, R., C. Hollier, T. Burks, and R. Frazier. 2015. First Report of Goss's Wilt of Corn Caused by Clavibacter michiganensis subsp. nebraskensis in Louisiana. Plant Dis. 99(9): 1268–1268.

St.Clair, D.A. 2010. Quantitative Disease Resistance and Quantitative Resistance Loci in Breeding. Annu. Rev. Phytopathol. 48(1): 247–268.

Storey, J.D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. 100(16): 9440–9445.

Strange, R.N., and P.R. Scott. 2005. Plant Disease: A Threat to Global Food Security. Annu. Rev. Phytopathol. 43(1): 83–116.

Suparyono, and J.K. Pataky. 1989. Influence of Host-Resistance and Growth Stage at the Time of Inoculation on Stewart Wilt and Goss Wilt Development and Sweet Corn Hybrid Yield. Plant Dis. 73(4): 339–345.

Suparyono, and J.K. Pataky. 1989a. Relationships between incidence and severity of Stewart's and Goss's bacterial wilts and yield of sweet corn hybrids. Crop Prot. 8:363–368.

Swarts, K., H. Li, R. Navarro, J. Alberto, D. An, M.C. Romay, S. Hearne, C. Acharya, J.C. Glaubitz, S. Mitchell, R.J. Elshire, E.S. Buckler, and P.J. Bradbury. 2014. Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants. Plant Genome 7(3)Available at https://dl.sciencesocieties.org/publications/tpg/abstracts/7/3/plantgenome2014.05.0023 (verified 1 February 2017).

Tambong, J.T., R. Xu, F. Daayf, S. Brière, G.J. Bilodeau, R. Tropiano, A. Hartke, L.M. Reid, M. Cott, T. Cote, and I. Agarkova. 2016. Genome Analysis and Development of a Multiplex TaqMan Real-Time PCR for Specific Identification and Detection of Clavibacter michiganensis subsp. nebraskensis. Phytopathology 106(12): 1473–1485.

Tanksley, S.D., and S.R. McCouch. 1997. Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. Science 277(5329): 1063–1066.

Tanksley, S.D., N.D. Young, A.H. Paterson, and M.W. Bonierbale. 1989. RFLP Mapping in Plant Breeding: New Tools for an Old Science. Nat. Biotechnol. 7(3): 257–264.

Technow, F., A. Bürger, and A.E. Melchinger. 2013. Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. G3 Bethesda Md 3(2): 197–203.

Treat, C.L., and W.F. Tracy. 1990. Inheritance of resistance to Goss's wilt in sweet corn. J Am Soc Hortic Sci 114(4): 672–674.

Treat, C.L., W.F. Tracy, P.N. Drolsom, and J.G. Coors. 1990. Inheritance of Resistance to Gosss Wilt in Maize. Crop Sci. 30(4): 893–896.

Ullstrup, A.J. 1972. The Impacts of the Southern Corn Leaf Blight Epidemics of 1970-1971. Annu. Rev. Phytopathol. 10(1): 37–50.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91(11): 4414–4423.

Vidaver, A.K., and M. Mandel. 1974. Corynebacterium-Nebraskense, a New, Orange-Pigmented Phytopathogenic Species. Int. J. Syst. Bacteriol. 24(4): 482–485.

Vidaver, A.K., D.C. Gross, D.S. Wysong, and B.L. Doupnik. 1981. Diversity of *Corynebacterium nebraskense* strains causing Goss's bacterial wilt and blight of corn. Plant Dis. 65:480–483.

Wallace, J.G., S.J. Larsson, and E.S. Buckler. 2014. Entering the second century of maize quantitative genetics. Heredity 112:30–38.

Wang, N., J.M. Akey, K. Zhang, R. Chakraborty, and L. Jin. 2002. Distribution of Recombination Crossovers and the Origin of Haplotype Blocks: The Interplay of Population History, Recombination, and Mutation. Am. J. Hum. Genet. 71(5): 1227–1234.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10(1): 57–63.

Webster, B.T. 2017. Factors contributing to the reemergence of Goss's wilt and blight of corn in the Upper Midwestern U.S. Available at http://conservancy.umn.edu/bitstream/handle/11299/185571/Webster_umn_0130M_17464.pdf?sequence=1&isAllowed=y.

Welz, H.G., and H.H. Geiger. 2000. Genes for resistance to northern corn leaf blight in diverse maize populations. Plant Breed. 119(1): 1–14.

Wisser, R.J., P.J. Balint-Kurti, and R.J. Nelson. 2006. The genetic architecture of disease resistance in maize: A synthesis of published studies. Phytopathology 96(2): 120–129.

Wisser, R.J., J.M. Kolkman, M.E. Patzoldt, J.B. Holland, J. Yu, M. Krakowsky, R.J. Nelson, and P.J. Balint-Kurti. 2011. Multivariate analysis of maize disease resistances suggests a pleiotropic genetic basis and implicates a GST gene. Proc. Natl. Acad. Sci. U. S. A. 108(18): 7339–7344.

Wolfe, M.D., I.Y. Rabbi, C. Egesi, M. Hamblin, R. Kawuki, P. Kulakow, R. Lozano, D.P.D. Carpio, P. Ramu, and J.-L. Jannink. 2016. Genome-Wide Association and Prediction Reveals Genetic Architecture of Cassava Mosaic Disease Resistance and Prospects for Rapid Genetic Improvement. Plant Genome 9(2)Available at https://dl.sciencesocieties.org/publications/tpg/abstracts/9/2/plantgenome2015.11.0118 (verified 5 January 2017).

Wray, N.R., J. Yang, B.J. Hayes, A.L. Price, M.E. Goddard, and P.M. Visscher. 2013. Pitfalls of predicting complex traits from SNPs. Nat. Rev. Genet. 14(7): 507–515.

WU, X., L. Xu, P. ZHAO, N. LI, L. WU, Y. HE, and S. WANG. 2015. Comparative transcriptome profiling of two maize near-isogenic lines differing in the allelic state for bacterial brown spot disease resistance. J. Integr. Agric. 14(4): 610–621.

Wysong, D.S., B. Doupnik, and L. Lane. 1981. Goss's wilt and corn lethal necrosis - can they become a major problem? Proc. Annu. Corn Sorghum Ind. Res. Conf. 36: 104–152.

Wysong, D.S., A.K. Vidaver, H. Stevens, and D. Stenberg. 1973. Occurrence and spread of an undescribed species of Corynebacterium pathogenic on corn in the western corn belt. Plant Dis. Report. 57: 291–294.

Yang, J., B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, P.A. Madden, A.C. Heath, N.G. Martin, G.W. Montgomery, M.E. Goddard, and P.M. Visscher. 2010a. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42(7): 565–569.

Yang, Q., G. Yin, Y. Guo, D. Zhang, S. Chen, and M. Xu. 2010b. A major QTL for resistance to Gibberella stalk rot in maize. Theor. Appl. Genet. 121(4): 673–687.

Yu, J., and E.S. Buckler. 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. 17(2): 155–160.

Yu, J., J.B. Holland, M.D. McMullen, and E.S. Buckler. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 178(1): 539–551.

Zhang, J., H. Zheng, Y. Li, H. Li, X. Liu, H. Qin, L. Dong, and D. Wang. 2016. Coexpression network analysis of the genes regulated by two types of resistance responses to powdery mildew in wheat. Sci. Rep. 6: srep23805.

Zheng, Z.-L., and Y. Zhao. 2013. Transcriptome comparison and gene coexpression network analysis provide a systems view of citrus response to "CandidatusLiberibacter asiaticus" infection. BMC Genomics 14: 27.

Zila, C.T. 2014. Traditional and genomic methods for improving Fusarium ear rot resistance in maize. Available at http://www.lib.ncsu.edu/resolver/1840.16/9195.

Zila, C.T., L.F. Samayoa, R. Santiago, A. Butrón, and J.B. Holland. 2013. A genome-wide association study reveals genes associated with fusarium ear rot resistance in a maize core diversity panel. G3 Bethesda Md 3(11): 2095–2104.